预测论基础

翁文波 著

石油工业出版社

预测论基础

翁 文 波 著

石 油 工 业 出 版 社

内 容 简 介

本书试图把自然科学和社会科学中某些预测问题在原则上统一起来，从而扩张了预测的领域，例如从微观粒子的性质到某些全球性工业的盛衰。本书在新的原则上提出超远程预测某些重要现象的可能性，例如预测一个城市在哪星期内下大暴雨或一个区域在哪月前后发生强烈地震的可能性。

预测论基础

翁文波著

*

石油工业出版社出版

（北京安定门外外馆东后街甲36号）

北京顺义燕华营印刷厂排版

曙光印刷厂印刷

新华书店北京发行所发行

*

850×1168毫米 32开本 $4^5/_8$印张 79千字 印1,501—6,180

1984年5月北京第1版 1984年8月北京第2次印刷

书号：15037·2502 精装定价：1.30元

科技新书目：77—133

FUNDAMENTALS OF FORECASTING THEORY

Weng Wen-Bo

(*Scientific Research Institute for Petroleum Exploration and Development, Beijing, China*)

Abstract

This thesis unifies the fundamentals of some forecasting problems in natural and social science, so that the domain of forecasting is extended from the physical property of microscopic particle to the life cycle of some world industry. The new principles point out the possibility of ultra longrange forecasting of some important phenomenon, such as the possibility of heavy downpour in a city within a certain week, or the possibility of some strong earthquake of a region around certain month.

目　录

绪 论

“凡事豫则立，不豫则废”和“人无远虑，必有近忧”等成语都说明预测的重要意义。现代技术预测的范围很广，包括社会预测（如人口等问题）、科技预测(如发展方向)、经济预测（如趋向、市场、管理等）和军事预测（如战略）等方面。

我国有些大专院校也开设技术预测（TF）课，但一般均隶属于社会科学中的技术经济范畴。本文希望在比较广泛的基础上，将技术经济领域内通用的预测技术和自然科学中的外推和预测方法尽可能有系统地汇合起来，组成“预测论”的一个基础概要。只是讨论内容一般是从静止、孤立、偏面、单纯和机械的体系出发，还只用到数代数系的模型，所以适用范围还会有很大的局限性。

预测可以区分为：(1)以统计学(STATISTICS)为基础、统计量（如平均值、方差等）为对象的统计预测，(2)以信息学（INFORMATICS）为基础、信号为对象的信息预测。本文主要讨论信息预测。

预测是带有主观性的。有人认为预测中的模式识别

是科学加艺术，所以，预测并不存在唯一的方法。

根据不同问题的性质，预测模型可以区分为确定模型和随机模型。本文主要讨论确定模型，最后二章也涉及随机性和有关问题。

第一章讨论信息的定义、信息过程和它的结果。第二章讨论预测体系 。由于预测体系是多种多样的。如：平稳的、不平稳的、线性的、非线性的、趋向的、周期性的、交变性的、均匀的、多重性的、突变或灾患性的、模糊的……等。本文未能作系统的分类和叙述，只是提出几种典型，希望能从典型扩张为类别。

从大量客观事物的观察中，可以归纳出许多规律，它们是对有关体系进行预测的基础。作为典型，本文讨论了几种有相当普遍规律的体系，它们是：对称体系；可数（量子化）体系，周而复始（周期性）体系等。其它信息体系，我们只讨论了两个极端情况，一个极端是事先就预定什么是（所要的）信息组成的体系。另一个极端是事先完全不规定什么是信息，而用相对简单的模型去拟合客观实际体系。在这两个根端之间，当然还有无数个中间类型。实际情况说明：上述典型信息，即：对称性、量子性和周期性等信息，常出现在各种信息体系之中。

回归可以看作是拟合模型中一种比较重要的特款，因为回归分析可以研究多体系的相关，所以可用于所谓

"因果预测"。在技术上，线性回归就是线性方程的最小二乘法拟合。当然，也可以用统计学的原则来定义，如一元正态线性回归，并用统计学的假设检查来判别。所以回归是从确定模型到随机模型的一个接界，也是信息预测和统计预测的一个汇合点。本文第七章有较详细的讨论。

最后，以随机体系的讨论作为一个转折，从统计预测的平均法（平滑法）发展到以否定随机性为原则的信息预测。

本文原来计划在一本有关"初级数据"一文出版以后再写的。但是由于预测技术的迅速发展，只好先写这本小册子。为了避免阅读上过分耗费时间，文字上力求简短，也不作详细的推导和讨论。而且由于预测论是一门新兴的学科，笔者也只能就几个简单的问题作一些定性的叙述，所引用的实例是一些示意性的单项预测，希望能引起读者的兴趣。笔者相信，读者如果有普速高等数学的一般知识和必要的耐心，一定能够理解笔者提出来的预测方法和方向，并且为开拓新的领域和途径作好准备，使将来的预测工作更有成效。

文中引用的实例，虽然原始数据是公开而真实的，但处理方法力求显明、简单，以便达到示意的目的。所以大部分结论带有习题性质，并非实际预测。

本文编写过程中，钱绍新同志提了有益的意见，吕牛领同志参加了整理定稿工作，特此致谢。

第一章 预测过程

统计学和信息学是两个不同的学科，在某些问题上，它们互相接界、互相汇合。统计预测和信息预测分别是以这两门学科为基础的，所以，根据习惯上的理解，它们各有独特的性质，但也并不排除它们之间的边缘接界或互相汇合。一种学科分类把统计学归属到信息学中，把统计量也作为信息的一类特款，这样，统计预测也就包含在信息预测之中了，那是另外一个问题，不在本文要讨论的范围之内。现在暂把统计预测和信息预测区分开来，并以信息预测作为主要研究对象。为此，首先谈什么是信息。

§1 信　　息

什么是信息？信息可以认为是信息体系中的元素、元素集或子体系。那么，什么是信息体系？本文认为信息体系是受人们主观定义约束的秩序类。主观定义的约束可以是：某种理解、信念、设想、定理、法则、规律、法律、契约、编码等。

有许多关于信息的定义。如："使消息中所描述事件

出现的不定性减少”、“消息中所含的意义，它不随载荷它的物理设备形式的改变而改变”[1]；又如：“信息这个名称的内容是我们对外界进行调节并使我们的调节为外界所了解时而与外界交换来的东西”[2]（维纳，N.Wiener，1894～1964），都和上述定义不相抵触。

自从申农（C.Shannon，1948）提出信息的概念并用概率来定义信息量、信息量的单位（BIT）及演算关系（加法）以来，信息学有了很大的发展。但是，现在不依据概率的信息定义受到非常广泛的注意[3]。信息的主观性是被维纳（《控制论》，1948）明确说明的。他把人、动物和机器的控制与通讯过程统一起来。本文以人们的主观定义为依据，动物和机器的控制，只有为人们所认识的条件下，才作为信息来处理。这和维纳的概念有原则上但无实际上的差异。不过本文用的“人们”一词和维纳的“人”有所不同。本文认为：单个人的某种信息思维，只有在传播到其它人并为其他人所共同理解的条件下才成为信息。“人们”之间的传播过程起到扩张、限制和中继等作用。这些作用又使信息超越了时间、空

①《英汉计算技术辞典》，人民邮电出版社，209～210页，1978。

②王鼎昌。“信息论的发展和意义”，自然辩证法通讯（2）1981，35页.

③H.W.Gottinger："Concepts and measures of information"，CISM，Courses and lectures No.29，Wien，New York，1975.

间、形式等的限制。在信息过程中，本文提出“介体”一词，介体包括主观的观点和客观的工具，其中也包括维纳的机器。

§2 信息交流

上文提到：人们之间的传播是信息形成的一项必要条件。在传播中，信息受到传播介体的扩张、限制和中继作用。开调查座谈会或专家小组会都是交流和形成信息的形式。在交流中，信息的扩张、限制和中继作用以不同方式表现出来。如有的人“最能说会道”、“最有威望”、“最有说服力”。结果，重要的信息可能在集体互相妥协中失真。为了信息保真，可以采用许多方法。例如“德尔菲(Delphi)技术”：由调查人员组织（可能匿名）专家组，用书面形式分别质询各专家的预测意见，再由调查人员综合、分析这些预测和现由，并向专家们提出一系列有关综合意见的问题。这样反复进行多次，可以得到最佳的或满意的结果。如果调查对象并非数目不大的专家而是广大群众，那末，随机抽查、随机蹬点也可能取得相对无偏的信息。

§3 几乎和可能

从客观体系中，以主观定义的信息来建立模型，存

在着主客观的矛盾。例如：从模型X中得出一个数值x和实际体系Z中观察到的对应数值$\hat{x}$，一般并不完全相同。

在确定模型中，x是一个确定值。如果认为差值（$\hat{x}-x$）是观察误差，而且大到不能令人满意，可能重新观察，即舍弃了$\hat{x}$；如果认为差值（$\hat{x}-x$）是模型X对体系Z的离差程度，而且大到不能令人满意，可能重建模型，即舍弃了x。在选择舍取观察值$\hat{x}$或模型值x前，有主观决定的临界标准，这个标准在线性规划中称为约束条件，在最优化过程中称为可行域边界。本文用符号ε表示这个可行临界值。$|\hat{x}-x|\leqslant\varepsilon$称为可行变程条件，它说明模型X对于体系Z的近似性。

在随机模型中，对应于一个x，模型给出一个随机变量的分布。x常常是这个分布的数学期望。从随机观点看问题，模型X是随机体系Z的形象。所以对应于一个x值，$\hat{x}$应该也是另一个随机变量的分布的数学期望。以上两种情况的随机变量的分布未必相同。对模型X或观察值$\hat{x}$的取舍也可以用与某一事先主观决定的置信水平相对应的置信区间作临界标准。如果沿用$|\hat{x}-x|\leqslant\varepsilon$的符号来表示置信区间条件，那末，相应的置信水平$(1-\alpha)$或$(1-\beta)$等可以说明模型X形象体系Z的可能性（概率）。

在某些事件预测中，如果实际的$\hat{x}$值和确定模型x值的近似性合乎主观要求，预测结论可以是："几乎如此"。如果实际的$\hat{x}$值和随机模型x值的可能性合乎主观要

求，预测结论可以是："可能如此"。

§4 体系和模型

一个体系可以看作是客观世界中被主观选取的一个局部，这个体系用符号 Z 来表示。为使主观选取的体系为一群人所共同认识，就得建立一个大家都能理解的模型，这个模型用符号 X 来表示。模型的建立是通过介体 Y 所产生的作用 O 而完成的。这个过程可用下式表示：

$$ZOY = X \quad (1)$$

在不同性质的介体 Y 的作用下，产生两类预测模型。一类是以体系 Z 中各元素的共性为基础的统计模型，因为体系中各元素的共性包涵在体系之中，在统计量定义以后，不再需要其它主观选择。从统计模型所产生的预测称为统计预测。另一类是以体系 Z 中各元素的特性为基础的信息模型。这种特性就是信息的定义。它也有二种基本类别，对应于两类信息模型。一类是以概率为基础的随机信息模型，申农提出的信息论就是用概率来定义信息的①。另一类是确定信息模型，信息的定义不涉及概率，而是由主观决定的。如"诧异价值"② 的概念

①Shannon.C E., Weaver.W："The mathematical theory of Communication" Illinois Univ.Press, 1949

②Longo .G ,"Information theory new trends and open problems" .Springer-Verlag, Wien-New York.12,1975.

把表示特性的信息定义为体系Z中各元素的差别。柯尔莫哥洛夫[①]等人提出的信息概念属于后一类。实际预测的信息模型可以同时包括这两类信息。从信息模型所产生的预测称为信息预测。

本文主要讨论信息预测的基础典型。

§5 映　　照

信息模型中，体系Z、介体Y和模型X的关系可参考映照关系作图解，如图1—1所示。图中直线Z代表体系；直线X代表模型；点Y代表介体；射线YO代表Z在X上的映照。在实际体系中，YO一般不是单值映照，Z中的元素和X中的元素可以但未必都是一一对应关系。实际上，一一对应关系是存在的，例如：一条消息符号链中的每一符号，和对应的正确代码是一一对应的。射线YO就表示代码转换。

显然，完全对应的映照只有在概念中存在，通常认为，在主观确定的可行临界值范围内偏离YO是可以被接受的。这一范围在确定模型中称为可行变程；在随机模型中称为置信区间，它是由置信水平决定的。

上文中的映照YO已超越了代数中函数的含义。例如

①Kolmogorov A.N, "Logical basis for information theory and probability theory" IEEE Trans. Information Theory, IT—14, 662, 1967.

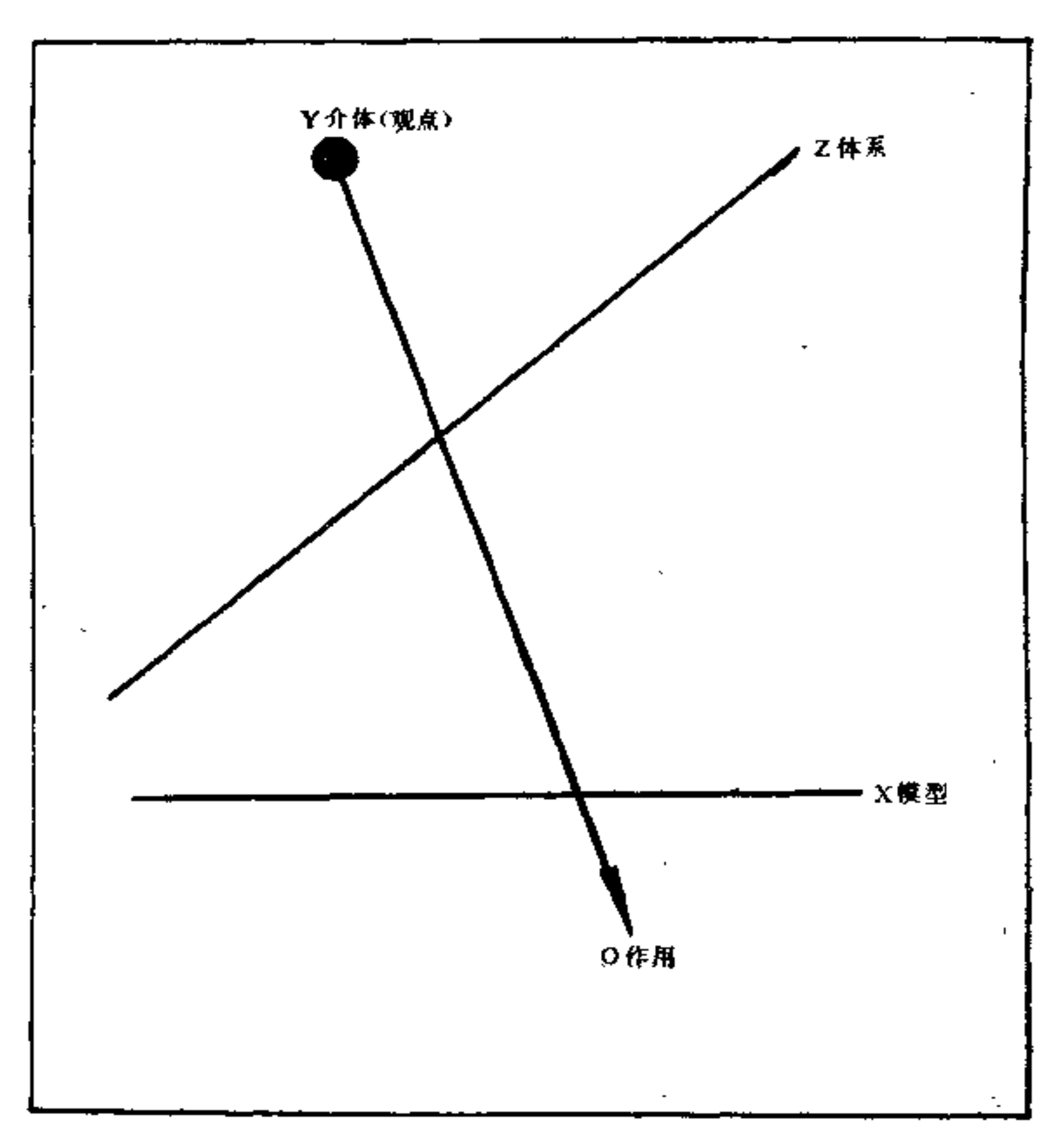


图1—1 信息模型

以地下地质构造为体系Z，通过地震勘探工程Y的实施O，得到磁带上数字序列集的模型X。从信息过程看，X中的数字序列集就是地下地质构造Z的信息映照。

§6 预测过程

预测过程包括从体系 Z_1 通过 Y_1O_1 映照到模型 X_1，再以模型 X_1 作为信源 Z_2，通过 Y_2O_2 映照到预测结论 X_2 二个阶段。在预测过程中，式(1)反复两次得到：

$$Z_1O_1Y_1 = X_1 \qquad X_1 = Z_2 \qquad Z_2O_2Y_2 = X_2 \qquad (2)$$

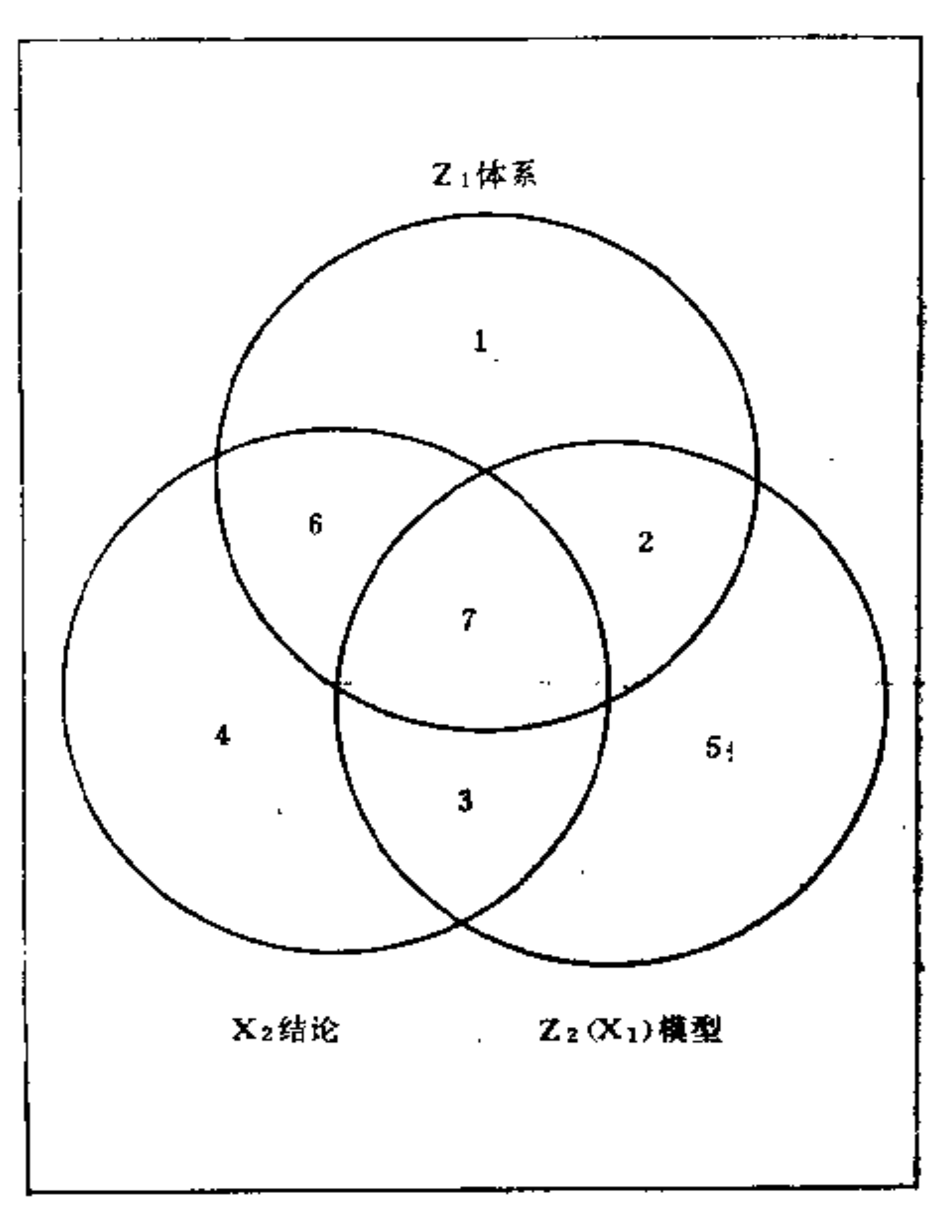


图 1—2 预测过程

式（2）可以图解为图1－2。图中三个圆圈分别代表体系Z_1，模型Z_2和结论X_2。圆圈中的面积表示主观定义的信息。和映照中的特款——函数相似，范氏（Venn）图可看作是图1—2的一种奇异特款。

Z_1、Z_2、X_2三个圆圈互相分割成七个区，表示了预测的七种情况，如表1—1所示。

表1—1　图1—2（预测过程）分区说明

区号	Z_1	Z_2	X_2	错误性质	
1	+			一次一型错误	漏失信息
2	+	+		二次一型错误	
3		+	+	一次二型错误	引入假信息
4			+	二次二型错误	
5		+		假正确（应预测而不预测）	
6	+		+	偶然正确（产生错觉）	
7	+	+	+		

区1是建模中漏失的信息，称为一次一型错误。区2是解模中漏失的信息，称为二次一型错误。如果区2是一个问题，那是一个正确的问题，区2未进入结论，就表示错误地解决了正确的问题。

区3是建模中引入的错误或假信息，称为一次二型错误。如果区3是一个问题，那是一个错误的问题，区3进入结论，就表示正确地解决了错误的问题。区4是解模中引入的错误或假信息，称为二次二型错误。

区5表示没有预测从错误模型中得出的结论。当然，不预测决不会出错，但应预测而不预测本身就是一种错误，所以区5代表假正确的错误。

区6表示错误地解决了一个错误的模型，却又碰巧成为正确。偶然正确不但不能扩张，并且产生错觉，所以区6也代表一种错误。

区7表示无错误，但并非完全、全部正确。只有当三个圆圈完全重合的情况下，才有完全、全部正确。

近年来，在预测失误的分析总结中，大多数人的注意力都集中在信息认识不当方面（如信息漏失和引入假信息）。实际上，预测的假正确和偶然正确也是应当设法避免的。

§7 反　　馈

前面说过，信息体系是人们主观定义的秩序类，所以信息不但有科学本身的问题，还有社会实践问题。例如，科学内部过去有过唯理和唯象的争议，现在，通过两者之间互相渗透互相补充，已经没有纯粹的唯理学派和唯象学派了。真理在争论中逐步澄清。又如，人们都要求改进，但改进的道路、标准和过程往往不同，需要在实践中不断检验、修改和完善，有时还会出现循环。总之，许多问题在体系设计（SYSTEM DESIGN）之前并没有暴露出来，一旦暴露应当及时解决，预测过程中的

信息反馈就是解决这些问题的方法之一。

信息的反馈，如结果检验，参量和模型的改变，子程序的重新组合等，可以预先作出反馈的设计，如图1—3中虚线长方框内 O_2 所起的作用。有些复杂的反馈过程，不能预先制定反馈程序的，就要用到虚线长方框以外，通过一定的途径（如人机对话等）使介体 Y_1 和 Y_2 发生作用，得到反馈。反馈不足，特别是动态反馈（DYNAMIC FEEDBACK）不足，可能引起预测的失误。

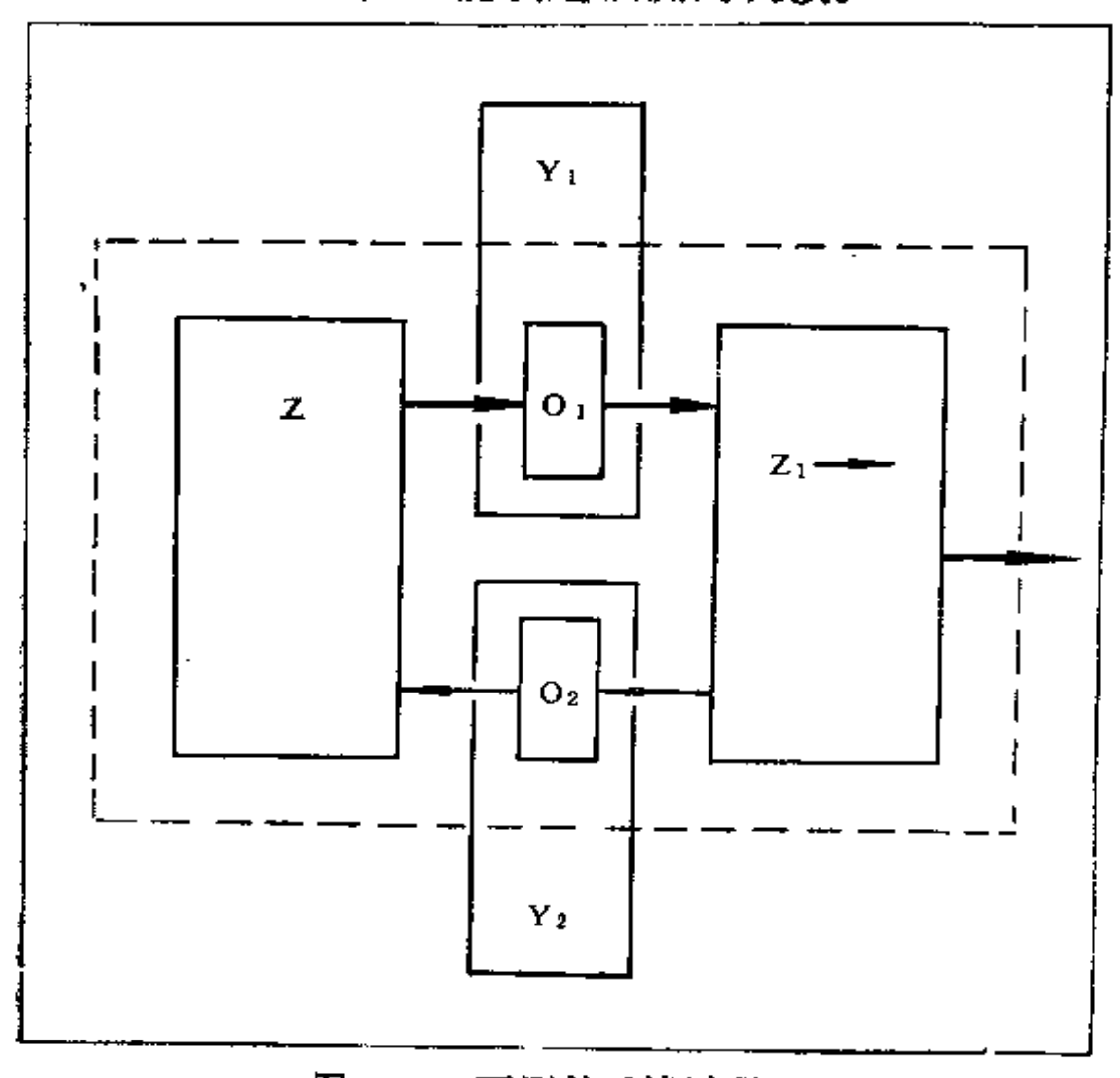


图1—3　预测的反馈过程

第二章　体系的属性分类

体系是客观世界的一个局部，这个局部的划分还没有一个固定的标准。本文叙述了几个典型的体系类别，在分析研究这些体系的过程中，它们显示出很不相同的特性。可以肯定，新的特性的类别还将继续被发现。

体系是预测的基础。没有搞清楚信息体系的特性（如没有考虑到长周期的变化）或者对体系的原始认识有了不同（如体系的范围起了变化，未考虑到的某种突然事件发生了等）都会引起预测的失误。动态反馈不足也是信息处理不当的一种形式，而不同体系的动态反馈可以有很大的差异。这些问题都说明研究体系属性的重要性。

§1 稳定体系

在确定模型中，一个封闭的、孤立的、不受外界影响的体系称为稳定体系。例如物理学中绝热、力平衡等体系。太阳系在天体演化过程中的一个相对短暂时间内，是一个力平衡的稳定体系。日、月蚀或海潮等的预测都是以体系的稳定性为主要条件的。

在随机模型中，分布函数不变的体系称为稳定体系。在均匀分布中，就是平均值不变。一个放射性物质放射某种粒子数属于这样一个体系。这个放射源在相对于半衰期很短的时间内，可以预测一段短时间内的粒子数。市场预测中的平均法也是以这类体系为前提的。移动平均法则假定体系相对稳定，也就是说变化很慢。正态分布体系的稳定条件，不但要求数学期望(平均值)不变，还要求方差不变。平衡随机过程是一种稳定体系，但认为过去的情况对未来情况的发生有着很强的影响。如纺纱过程中，一束纱通过某一定点时的截面积与前一个短时间该束纱通过该点的截面积有关，时间越短关系越密切。这就发展为指数移动平均或指数平滑等预测方法。所谓“指数”就是衡量这种关系的一种参量。马尔科夫过程或者说无后效的随机过程(包括链)，特别是在多而有限的不同状态之间，一定时间间隔内状态转移不变的时齐的马尔科夫过程也是一种稳定体系。在转移概率可以估计的条件下，马尔科夫链被用于预测升学、失业、不同牌号的同类商品的销售比例等问题。

§2 计量体系

在社会科学范畴内的技术经济预测模型中，有一类称为计量模型。这主要是对立于判断法预测的一种分

类。判断法包括：历史比拟、市场定点探测、开专家或顾客座谈会，征集意见等直观方法。这类直观方法一般不能用数学作为信息的中继体。计量模型是一个新近发展的预测模型，有时称为正则的预测模型，主要特征是以数字作为信息的中继体。也有人把计量模型局限于带信息反馈或动态反馈等一类模型而把时间序列（外延法）和回归模型排除在外。

在自然科学范畴内，技术经济中的计量模型大体对应于以数字为中继体的数代数和数组代数模型。而几何模型和集合代数、逻辑代数、符号代数模型等基本被排除在外。

§3 复合体系

一个可以产生多种不同信息的体系可称为复合体系。不同种信息可以来自：性质和类型不同或形式不同的种种信息渠道，也可以来自带有多种信息的一个渠道。

比较单纯的复合体系是简单的迭加体系。设$\langle x_i \rangle$表示容量为n的序列：

$$\langle x_i \rangle = \langle x_1, x_2, x_3, \cdots, x_i, \cdots, x_n \rangle$$

其中x_i都是数值，i是自然数构成的下标。符号$\langle\ \rangle$表示其中的数值是有序的，即$\langle 1, 2 \rangle \neq \langle 2, 1 \rangle$

如果序列$\langle x_i \rangle$是由序列$\langle f_i \rangle$，$\langle g_i \rangle$，…等子序列

相加而成的：

$$\langle x_i \rangle = \langle f_i \rangle + \langle g_i \rangle + \cdots$$

称为迭加体系，它们的k阶差分也有迭加关系：

$$\nabla^k \langle x_i \rangle = \nabla^k \langle f_i \rangle + \nabla^k \langle g_i \rangle + \cdots$$

一般的复合体系未必是由体系划分的全部子体系组合而成的。我们避开了划分的严格条件，笼统地用"分解"一词说明一个复合体系的组成。

分解演算可以用加法，如：

$$\langle x_i \rangle = \langle f_i \rangle + \langle g_i \rangle + \cdots$$

也可以用乘法，如：

$$\langle x_i \rangle = \langle f_i \rangle \cdot \langle g_i \rangle \cdots\cdots$$

一种特款是把序列 $\langle x_i \rangle$ 分解为长期趋向序列 $f\langle x_i \rangle$（如直线型序列或抛物线型序列等）和一个零和序列 $\langle g_i \rangle$（即 $\sum_{i=1}^{n} g_i = 0$）之和。

零和序列还可以分解成各种半序集[①]，办法之一是建立一个包含零和序列 $\langle g_i \rangle$ 的连续函数 $g(t)$，再以某种条件建立半序集。如：

$$\langle t_j \rangle = \langle t_j \mid g(t) = 0 \rangle$$

$$\langle t_j \rangle = \langle t_j \mid g(t) = a \rangle \qquad a\text{为任意数}$$

$$\langle t_j \rangle = \langle t_j \mid \frac{dg}{dt} = 0 \rangle$$

① 容量为n的半序集是指：

$\langle x_i \rangle = \langle x_1, x_2, \cdots, x_i, \cdots, x_n \rangle$

其中 $x_1 \leqslant x_2 \leqslant \cdots \leqslant x_i \leqslant \cdots \leqslant x_n$

或 $x_1 \geqslant x_2 \geqslant \cdots \geqslant x_i \geqslant \cdots \geqslant x_n$

$$\langle t_i \rangle = \langle t_i \mid \frac{d^2 g}{dt^2} = 0 \rangle$$

其中第一种又可以称为符号位半序集。符号位技术已被应用于某些地震勘探工程中。

比较基本的复合体系是多重体系。多重体系在本质上有一个以上完全不同的属性。例如微观粒子就具有波粒两重性。下面我们也提出一个多重性体系的例子。

多重性体系例——原子核幻数模型

从两种观点对原子核幻数：2、20（6+14）、28、50、82、126建立模型。

第一种观点认为幻数从属于物质，必然占有空间，可以表示体积，猜想它有三次方程形式，如：

$$x_{i1} = 1 + \left(1 + \frac{2}{3} i\right)^3 \qquad i=0,\ 1,\ 2,\ \cdots$$

另一种观点认为幻数是一种量子数或者和量子数有关，必然是自然数，至少也是$\frac{1}{2}$的倍数。取自然数建立模型，得到：

$$x_{i2} = \frac{1}{3}(i+1)(i^2+2i+6) \qquad i=0,\ 1,\ 2,\ \cdots$$

由此可以导出：

$$(x_{i2} - x_{i1}) = i(i-3)(i-6)/27$$

所以：

$$(x_{i2} - x_{i1})_{i=1} + (x_{i2} - x_{i1})_{i=5} = 0$$

$(x_{12}-x_{11})_{i=2}+(x_{12}-x_{11})_{i=4}=0$

x_{11}和x_{12}值的比较如表2—1。

表2—1 x_{11}和x_{12}的值比较

i	0	1	2	3	4	5	6	7
x_{11}	2	～5.63	～13.70	28	～50.30	～82.37	126	～182.96
x_{12}	2	6	14	28	50	82	126	184
$x_{12}-x_{11}$	0	+0.37	+0.30	0	−0.30	−0.37	0	+1.04

由上表和上述关系可以预测：如果两种观点各代表两个核稳定条件，在 $i<7$ 域内 $|x_{12}-x_{11}|\leqslant 10/27$ 两个条件都接近满足。如果 $i\geqslant 7$，$x_{12}\geqslant 184$，已失去幻数意义。

§4 突变体系

某些数学工具(如不连续函数)可以表达数学模型中自变量变化很小、应变量变化很大的状态。电脉冲、爆炸、结晶、破裂、碰撞、油层水淹等现象处于这种状态。经济危机、企业破产等现象也处于这种状态。

在相对简单的模型中，函数的不连续点可以形象突变现象。等轴双曲线函数，正切函数等都有这类不连续点。

狄拉克函数具备突变的性质

$$x(t)=\begin{cases}\infty & t=0\\ 0 & t\neq 0\end{cases}$$

沃西函数也具备这类性质，如沃西一阶函数：

$$x(t)=\begin{cases}1 & 0<t<\frac{1}{2}\\ 0 & t=0,\ \frac{1}{2},\ 1\\ -1 & \frac{1}{2}<t<1\end{cases}$$

$$x(t)=x(t+I)=-x(t+\frac{I}{2}) \qquad I\ 为整数$$

沃西函数多项式可以和傅里叶级数相比拟，其特点是便于表达突变状态，算计上占时间较少。如果为了形象单纯的零和时间序列中的周期性，本来可以不用三角函数，但沃西函数使用不方便，并且内涵信息较少。例如在 x-t 座标上的西个点，三角函数拟合中可能有一个频率信号，而在沃西函数拟合中未必有这样的可能信号。

一种小数函数兼有这西种函数的某些特性，如：

$$x(t)=\begin{cases}t-\frac{1}{2} & 0<t<1\\ 0 & t=0,\ 1,\ -1\\ \frac{1}{2}+t & -1<t<0\end{cases}$$

$$x(t)=x(t+I)=x(t+\frac{I}{2})-\frac{1}{2} \qquad I\ 为整数$$

1972年灾患理论提出用几种拓扑面来形象这类客观

世界中突然发生的事件①。也就是说用拓扑面在三维空间中的位置作为突变事件的模型②。这类曲面可以用三次方到六次方的曲面函数来表示，体系的发展被形象为受到垂直向下或向上趋向所控制的曲面上的质点。当质点走到曲面拐弯处时会突然落到曲面下部的另一个部位上。这类模型在一定程度上承认在突变事件发生前已有某些迹象表明突变的趋向。灾患理论的模型可用于各种对立状态突然转变的过程中，例如结晶和溶解，疾病的暴发和受控制，暴雨和地震的发生，动物行为中的进攻和退

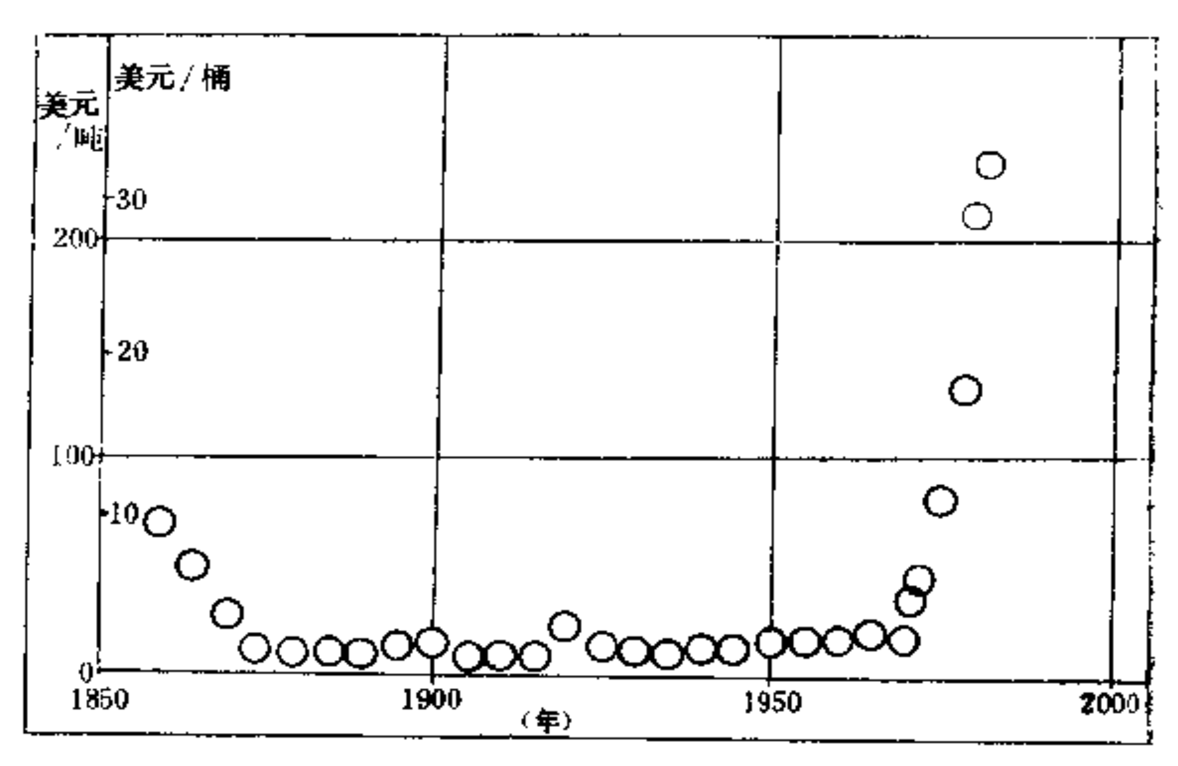


图2—1 世界石油价格

① René Thom, "Stabilité structurelle et morphogénèse", 1972.

② Zeeman E. C., "Catastrophe theory", selected papers, 1972—1977, Addison-Wesley Publishing Co. London, 1977

却，物价的暴涨和暴落等。

突变体系例一——世界石油价格

世界石油价格的暴涨，近似一种突变事件，如图2—1所示。1960年8月8日，产油国税收被削减，触发1973年10月6日世界石油危机[①]。

§5 动态体系

动态体系是由几个互相有关并有时差的子体系组成。一个简单的动态体系可由三个子体系：$A(t+t_a)$、$B(t+t_b)$、$C(t+t_c)$所组成，式中t为时间变量，t_a、t_b、t_c为时差参量。A、B、C的相互关系可用线性公式表示：

$$A=a_0+a_2B+a_3C$$

$$B=b_0+b_1A+b_3C$$

$$C=c_0+c_1A+c_2B$$

用矩阵表示为：

$$\begin{pmatrix} 1 & -a_2 & -a_3 \\ -b_1 & 1 & -b_3 \\ -c_1 & -c_2 & 1 \end{pmatrix}\begin{pmatrix} A \\ B \\ C \end{pmatrix}=\begin{pmatrix} a_0 \\ b_0 \\ c_0 \end{pmatrix}$$

这个体系存在的条件是：

1）因为A、B、C都是在有限的世界（如地球）上存在，所以：

---

①Alvin Toffler, "The third wave", 1980.

$$\begin{vmatrix} 1 & -a_2 & -a_3 \\ -b_1 & 1 & -b_3 \\ -c_1 & -c_2 & 1 \end{vmatrix} \neq 0$$

2）因为A、B、C都是客观存在的实体，所以：

$$A>0,\ B>0,\ C>0$$

3）因为A、B、C存在于一个有限体系中，所以有约束条件：

$$A\leqslant A_L,\ B\leqslant B_L,\ C\leqslant C_L$$

式中 A_L,B_L,C_L 分别代表上限。A、B、C 等子体系可以是世界上的人口、能源消费、环境质量；也可以是一个工厂的订货单、人事、库存、积压货单、生产率等；也可以是一项国际契约的国家主权、地区产值、资源利用、工业化程度等；也可以是一个政府的国民生产总值、利率、税收、投资、消费、工资率等①。

动态体系中各子体系的模型较难建立，有关参量如 a_0，b_0，… 等也不易估计，而且，它们之间的互相作用可能使体系模型不稳定。所以可能会导出很有争论的结论。例如世界人民生活水平是否会持续下降等。

§6 互逆体系

逆就是相反。如果把预测信息过程写成映照形式：

① Forrester J. W.: Industrial dynamics 1961, Urban dynamics 1969, World dynamics 1973.

$$Z_1 \xrightarrow{O_1} X_1 \xrightarrow{O_2} X_2$$

它的逆过程就是：

$$X_2 \xrightarrow{O_2^{-1}} X_1 \xrightarrow{O_1^{-1}} Z_1$$

如果O_1和O_2都是单值映照（bi－jective）即有：

$$(O_1O_2)^{-1}=O_2^{-1}\quad O_1^{-1}$$

信息体系中的信息元素x，单位信息元素1，和逆信息元素x^{-1}的关系是：

$$x\bullet x^{-1}=1$$

式中符号“•”表示合成；指数“－1”表示逆或相反，“＝”表示合成结果，“1”表示信息贫乏。

如果把互逆信息之一反过来，再合或就可得到信息更为丰富的结果。

$$x\bullet(x^{-1})^{-1}>x$$
$$x>1$$

式中符号“＞”表示更为丰富或加强。

在实际信息体系中，经济繁荣和衰退；物价暴涨和暴落，年景的干旱和水涝都可看作是互逆信息。

互逆体系例一——太阳黑子活动预测

用太阳黑子活动的极大年和极小年资料，分别建立两个周期函数多项式，再将极小年模型多项式各项正负反一下加在极大年模型多项式上得出综合模型，预测结

果见表2—2。从表中可以看出：在许多预测区间出现二个预测值，极大年的第一个预测值比较接近实际；极小年的第二个预测值比较接近实际。由此可见这个模型还没有揭露出太阳黑子活动规律的全部主要信息。

表2—2 太阳黑子活动极大、极小年预测

极大年			极小年		
预测值		实际值	预测值		实际值
1806.0		1805.2		1811.5	1810.6
1816.5、	1819.5	1816.4		1821.5	1823.3
1828.0、	1830.0	1829.9	1832.0、	1834.0	1833.9
1839.5、	1840.5	1837.2	1843.0、	1844.5	1843.5
1848.5、	1850.5	1848.1	1854.5、	1855.5	1856.0
1860.0、		1860.1	1865.5、	1867.5	1867.2
1870.5、	1873.5	1870.9		1889.0	1889.6
1884.0、		1883.9		1889.0	1889.6
1894.5、	1897.5	1894.1	1900.5、	1901.5	1901.7
1904.5、	1907.0	1907.0		1912.0	1913.6
1914.5、	1917.5	1917.6	1922.0、	1923.5	1823 6
1928.0、	1929.0	1928.4		1932.5	1933 8
1937.5、	1938.5	1937.4	1942.5、	1944.5	1944 2
1948.5、		1947.5	1953.5、	1955.5	1954 3
1958.0、		1957.9	1960.0、	1966.0	1964 8
1968.5、	1972.0	1968.9		1976.0	
1982.0、			1986.0、	1987.0	
1992.0、				2001.5	
2006.5、			2010.0、	2013.0	
2016.5、	2018.0		2021.5、	2023.5	
2026.5、					

§7 模糊体系

如果原始体系Z_1或观察到的形象X_1原本是模糊的[①]，那末用清晰模型概括自然现象可能一筹莫展或导出不完全真实或不完整的结果。模糊体系的基本概念是[②]：一些复杂的实际问题，不适当地要求准确和明确的解成为困难，那末，应当用描述和分析的方法，来适应那些不准确的知识分界，适应我们对价值的判断和成绩的评价中的主现性。例如：天体对地球水圈（或层）的潮汐作用是比较清楚的，但天体对大气圈（或层）的作用还是模糊的。这个问题我国科学家已研究多年，还是没有能提出一套严格的定量预报模型。如果用模糊的概念处理，还是可以得出虽然模糊、但还有参考价值的预测结果（见例一）。同类的体系亦称："灰色体系"。

模糊体系例一——北京暴雨和下弦月

周万福提出了一系列地区下暴雨的日子和近地天体位置模糊相关。从他的一份资料中看出：北京日降水量

---

① L.A.Zadeh, Outline of a new approach to the anaslyis of complex systems and decision processes IEEE trans. Systems, Man, and Cybernetics Vol SMC-3 Jan 1973.

② Mamdani, E.H., Report on the 2nd round table discussion of fuzzy automata and dicision process, Proc. 6th IFAC Congress, Boston, 1975.

在100毫米以上的大暴雨日子几乎有一半在月相为下弦的农历二十三日到农历二十七日这五天之间。我们用其它预测方法已初步估计到1983年8月初旬到中旬将下大暴雨，但还不易确定具体日子，查历本知道1983年8月1日～1983年8月5日正是农历二十三日到二十七日。因此可预报这几天内下暴雨。实际情况是：1983年8月4日～1981年8月5日下暴雨，降水量超过70毫米，局部地区超过260毫米。

§8 不定体系

信息的一种定义是："使消息中所描述的事件出现的不定性减少。若不提供信息，不定性会大一些。"经过信息的作用，不定性如果已不存在，那么预测问题也就不需要了。所以，对于预测体系，不定性总是存在的。

信息体系的模型可以是不定方程。不定方程的定义是："对子 n 个未知数、m 个整系数代数方程，如果 $m<n$ 且有解，那么有无限多个解。如果在无限多个解当中，只想求出所有整数解，那么，这样的问题叫数解不定方程问题，有关的方程叫数不定方程"。[1] 希腊人丢番图（Diophantus, 250 A.D.）首先提出并解决了求整数解问题。这类方程又称为丢番图方程。这个定义的第一部分提

① 北京教育学院师范教研室，《整数基础知识》，122页，1982。

出无限多个解，对信息体系来说，就是预测可以是无限的。定义的第二部分提出整数解。本章多重体系例一中有两个方程，其中任何一个都有无限多个解，但两个方程组成的“几乎”联立方程（只要求在某种精度范围内求解）只能有有限个解。第四章讨论的I信息系就是求整数解，其中最重要的特款是可公度信息系。还可以把整数进一步限制为自然数（这可能也是丢番图的原意），第四章中的Z信息系就是求自然数解。

用不定方程来描述不定信息系只是一种比拟。不定方程的必要条件是 $m<n$，而不定信息体系的条件则还要广泛一些。

第三章 对称和守恒

照字面上讲，对称就是两个东西相对又相称，或者说相仿、相等。因此把这两个东西对换一下，好象没有动过一样[1]。还有一种说法认为当某种变动使有些东西不变时，即有对称存在。对称是自然界很普遍的现象。矿物的晶体、生物的形态都表示出对称，人们也创造许多对称的产品。凡是对称的形象、符号或实物经过空间平移、旋转又可恢复到原来的样子，所以对称又联系到守恒。在近代物理的概念中，有许多对称演算和有关守恒原则的例子[2]。对称和守恒是密切联系着的概念。它们在客观世界中是这样普遍，使人们可以在对某些体系还没有完全理解以前就作出对称或守恒的预测。

对称的概念和抽象代数中"群"的概念有关。"群"的概念有助于建立重子分类的八度法模型（盖尔曼和尼曼，1961～1962），并且预测："失踪了的"第10个粒子应具有这样的性质：电荷Q＝－1，奇异数S＝－3，质量为1680兆电子伏特，自旋3／2，宇称十。两年后（1964年

① 段学复，《对称》，人民教育出版社，5页，1974.

② Arthur Beiser, Concepts of modern physics, Mcgraw-Hill Book Co., p448, 1973.

初），在布鲁克海汶实验室发现了这个Ω^-粒子，正好符合上述预测。"群"的概念也有助于解开魔方。

尽管世界每时每刻都在变化着，但在一定时间、空间内，某些局部（体系）的某些因子变化很小，可以暂时认为是不变的。那末，守恒原则就是某些体系预测的基础。

近年来经济技术预测中最重要的两条通用原则都是以守恒原则为基础的。它们是①：

1)连贯原则：是说未来有些象过去的样子。

2)类推原则：是说类似体系的结构和变化具有一定的模式。

§1 对　　称

概括地说，对称是一种关系。以二元关系为例，设有序组$\langle x,y\rangle$，如果对任意$\langle x,y\rangle\in R$都能推出$\langle y,x\rangle\in R$，那么R是对称关系。如果$\langle x,y\rangle\in R$，又$\langle y,x\rangle\in R$，必有$x=y$，则R为反对称关系。如果$\langle x,y\rangle\in R$，可能有$\langle y,x\rangle\in R$，那么为不对称关系。对称关系是等价关系的一个重要的必要条件。

在演算关系中，交换律是对称关系，加法（+）和

① 朱景尧，《统计研究》第二辑、中国财政经济出版社，54—55页1981。

乘法（×）都服从交换律，所以有对称关系。减法（－）和除法（÷）不服从交换律，它们有反对称关系。在一个有恒等元素的独异半群(monoid)中，如果对于元素x，存在着另一个元素 x′，使它们在半群演算关系中合成为恒等元素，那么 x 和 x′是互逆元素。如在整数加法独异半群（I，＋，O）中，I 是整数，演算关系是加法，恒等元素为O（因为 I+O=I）。则 I 和（－I）是互逆元素。因为〈I，（－I）〉和〈（－I），I〉有零和（其和为零）的关系。它们是对称的。同样，在整数乘法独异半群（I，X，1）中，I 是整数，演算关系是乘法，恒等元素为 1（因为 I×1 =I）。则I和（$\frac{1}{I}$）是互逆元素。I和（$\frac{1}{I}$）在这个半群中对称。

在对称函数中，如 $\psi(x)=\psi(-x)$，$\psi(x)$称为偶函数，或宇称 P 为正，即 P ＝＋1，可表示为：$\psi(x)=P\psi(-x)$；在反对称函数中，如$\psi(x)=-\psi(-x)$，$\psi(x)$称为奇函数，或宇称为负，即P=－1，仍可表示为：$\psi(x)=P\psi(-x)$。

§2 对称多项式

对于容量为 n 的数据序列〈x(t)〉，可以建立一个 n 次多项式模型：

$$x(t)=a_nt^n+a_{n-1}t^{n-1}+\cdots+a_1t+a_0$$

在复数域内可以用因子分解定理分解为：

$$x(t)=a_n(t-t_1)(t-t_2)\cdots(t-t_n)$$

其中t_1，t_2，…，t_n是方程$x(t)=0$的根。

韦达公式给出：

$$t_1+t_2+\cdots+t_n=-\frac{a_{n-1}}{a_n}$$

$$t_1t_2+\cdots+t_1t_n+t_2t_3+\cdots+t_2t_n+\cdots+t_{n-1}t_n=\frac{a_{n-2}}{a_n}$$

……

$$t_1t_2\cdots t_n=(-1)^n\frac{a_0}{a_n}$$

在上面的公式中，不论怎样交换两个t_i，或者任意排列t_1，t_2，…，t_n，$x(t)$都不会改变，这样的多项式称为初级对称多项式。当然，根集$\{t_1, t_2, \cdots, t_n\}$中的元素已经不是原始数据序列中的元素了。

对称多项式可以进一步扩张到排列群的概念中。

§3 投入产出守恒

投入产出平衡是投入产出法预测的基础。达是一般账目上收支平衡的扩张。近年来，投入产出预测方法已经为某些计划、经济部门所采用。

设有投入产出体系。体系中的元素用带有二个下标的符号x_{ij}来代表，其中第一个下标i为产出指标，第二

个下标 j 为投入指标，x_{ij} 表示由元素 i 产出并投入元素 j 的量。这个体系可以划分为内部子体系和外部子体系。内部子体系可以是一个国家、一个部、一个企业在一定时期内货币价值、能源、物资等的流动。内部子体系的元素下标用1到n的正整数来表示。外部子体系可以是国家积累 国家消费、出口、税金、国民收入等。外部子体系的元素下标用0来表示。

内部子体系的流动可以从投入和产出两个方面分析。从产出方面看，因为 x_{ij} 是 i 元素的产出投入到 j 元素中去的量，所以元素 i 的内部总产出为：

$$\sum_{j=1}^{n} x_{ij} \qquad \text{（元素 } i \text{ 的内部总产出）}$$

同样，从投入方面看，从内部子体系的全部元素的产出投入到元素 j 中即为元素 j 的内部总投入：

$$\sum_{i=1}^{n} x_{ij} \qquad \text{（元素 } j \text{ 的内部总投入）}$$

在内部子体系中，全部总产出等于全部总投入，所以有：

1）守恒原则一：

$$\sum_{i=1}^{n}\sum_{j=1}^{n} x_{ij}=\sum_{j=1}^{n}\sum_{i=1}^{n} x_{ij}$$

再考虑外部子体系，它的流动也可以从投入和产出两方面来分析，令 x_{i0} 表示从元素 i 产出向外部体系投

入，x_{0j}表示从外部体系向元素j投入。在国家工农业经济体系中，外部体系向各元素的总投入称为国民总收入。即：

$$\sum_{j=1}^{n} x_{0j}$$ （国民总收入）

从产出方面看，元素i投向内部和外部子体系（即全体系）的总产出是：

$$\sum_{j=0}^{n} x_{ij}$$ （元素i总产出）

从投入方面看，全体系向元素j的总投入为：

$$\sum_{i=0}^{n} x_{ij}$$ （元素j的总收入）

因为各元素的总产出等于该元素的总投入，所以有：

2）守恒原则二

$$X_1=\sum_{j=0}^{n} x_{1j}=\sum_{i=0}^{n} x_{i1}$$ （1号元素平衡）

$$X_2=\sum_{j=0} x_{2j}=\sum_{i=0}^{n} x_{i2}$$ （2号元素平衡）

………

$$X_n=\sum_{j=0}^{n} x_{nj}=\sum_{i=0}^{n} x_{in}$$ （n号元素平衡）

将上述公式相加，就得到：

3）守恒原则三：

$$\sum_{k=1}^{n}\sum_{j=0}^{n}x_{kj}=\sum_{k=1}^{n}\sum_{i=0}^{n}x_{in} \qquad \text{（体系总平衡）}$$

在国家工农业经济体系中，上述值称为“国民经济总产值”。

对某一元素 k 来说，从元素 i 产出向元素k投入的是 x_{ik}，而这个元素的总投入为：

$$\sum_{i=0}^{n}x_{ik}=X_k$$

两者的比值是：

$$a_{ik}=\frac{x_{ik}}{X_k}=\frac{x_{ik}}{\sum_{i=0}^{n}x_{ik}}$$

a 称为元素 k 消耗元素 i 产出的“直接消耗系数”或称“消耗定额”。由上式可知：

$$a_{ik}\leqslant 1$$

从直接消耗系数 a_{ik} 可以导出完全消耗系数 b_{ik}，它们之间的关系如下：

令：

$$A=\begin{pmatrix}(1-a_{11}) & -a_{12} & \cdots & -a_{1k} & \cdots & -a_{1n}\\ -a_{21} & (1-a_{22}) & \cdots & -a_{2k} & \cdots & -a_{2n}\\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots\\ -a_{k1} & -a_{k2} & \cdots & (1-a_{kk}) & \cdots & -a_{kn}\\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots\\ -a_{n1} & -a_{n2} & \cdots & -a_{nk} & \cdots & (1-a_{nn})\end{pmatrix}$$

$$B=\begin{pmatrix} (1+b_{11}) & b_{12} & \cdots & b_{1k} & \cdots & b_{1n} \\ b_{21} & (1+b_{22}) & \cdots & b_{2k} & \cdots & b_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ b_{k1} & b_{k2} & \cdots & (1+b_{kk}) & \cdots & b_{kn} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ b_{n1} & b_{n2} & \cdots & b_{nk} & \cdots & (1+b_{nn}) \end{pmatrix}$$

矩阵A和B有互逆关系：

$$A^{-1}=B$$

它们和有关向量的关系有：

$$A\cdot\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_k \\ \vdots \\ X_n \end{pmatrix}=\begin{pmatrix} x_{10} \\ x_{20} \\ \vdots \\ x_{k0} \\ \vdots \\ x_{n0} \end{pmatrix},\quad B\cdot\begin{pmatrix} x_{10} \\ x_{20} \\ \vdots \\ x_{k0} \\ \vdots \\ x_{n0} \end{pmatrix}=\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_k \\ \vdots \\ X_n \end{pmatrix}$$

从Leontief① 建立的一次世界性经济投入产出模型可以看出，分析结果受到单项预测（如裁军预测）的影响。

§4 标度模拟

标度模拟通称模拟，它是相似原理的扩张。相似原理以守恒为根据。例如相似三角形中三个角的角度守

① Wassily W. Leontief, Scientific American 243 (3), 1980 (Scpt).

恒，三条边的比例守恒。从相似三角形（homothetic triangle）又扩张到同位相似二次曲线、同位相似曲线、同位相似图形、同位相似变换等。

在十九世纪，标度模拟方法被用来解决分式分析（fractional analysis）问题。所谓分式分析就是：在一个问题无法（或不需要）求得完全的解时，取得有关答案的一部份信息的过程，也就是说，即使不能求得完全准确的解，也要得到尽可能多的信息。

在力学中应用标度模拟的主要方法是：量纲分析（π理论）和力的无量纲比值。根据力学的模似原则，两个力学体系之间，如果存在着几何和力学的相似性，就有相似的性质和解。如果两个力学体系可以变换成为几何和力学的相似体系，也有相似的性质和解。流体力学的标度模型就是根据这一原则建立起来的，如用于预测飞行器空气动力学性能的风洞实验，又如用于预测船只和河道水动力学性能的水池实验等。十九世纪以来，物理学界注意到重力、电力、磁力等系统之间具有相似性，这就产生了一种信念：几种力学系统不仅相似，还有可能统一起来。

在力学范围以外，标度模拟的应用还很少见，例如用瓶养果蝇或植物群落来模似地球上的人口增长并不符合标度模拟的原则。不过，现代的信息论研究中，已经有一些人想把力学、电磁学、热学和经济学建立一个统

一的模型[1]。

§5 类　比

从守恒扩张到相似（标度模似），又扩张到类比。一般讲，类比不再局限于定量的比例关系。类比是从两个体系中已经确定的互相类似的性质，预测尚未确知的互相类似的性质。例如，两种体系成因相似，体系性质可能也相似。类比预测的应用范围较广泛，如：

1）排演预测　对于预期可能发生的事物，有时用实体模拟来预测将来可能发生的情况。这类排演预测一般用于比较复杂的体达不到标度模拟的要求的体系。舞台彩排是一种对实际演出的预测；军事演习是一种对可能发生的军事事件的预测。其中一部分技术已逐步发展为用符号演算未部分代替实体模拟。

2）先驱预测　某体系的发展过程，可能找出一个可类比的体系的发展过程作为预测的依据。例如，民用飞机的速度是随时间增加的，发展的趋势和军用飞机速度的发展趋势平行，并可找出时间上的延迟程度。军用飞机的速度又以世界上飞机速度比赛的最高记录为先驱。民用汽车的速度也以体育比赛中跑车速度的最高记

① Evans F.T., Physics, Structure and Information, ref. G. Longo, Information theory, new trends and open problems, Splinger-Verlag, Wien–N.Y.p61, 1975。

录为先驱。

类比预测例一——从地球的对称预测油田

地球的形状大体对称于地心，地球的自转运动又轴对称于地轴。这种对称性使太平洋东西西边纬度相近的地带上的含油气盆地显示出一定的近似性。笔者在1958年《地球形态的发展》一文中曾经预测："……在较高纬度带，如相当于加拿大落矶山油区（N50°）与我国东北（N45°），沉积上也有一些相似之处，加下古生界地层变质；泥盆系石炭系地层存在；二迭系三迭系地层发育较差，西白垩系则发育很好。在加拿大落矶山油区，油层就在泥盆系、石炭系与白垩系地层中。"1959年发现大庆油田，其主要产层也属白垩系。当时还把华北地区和北美粉江盆地类比，指出该盆地的含油层有：奥陶系、石炭系、二迭系、三迭系、白垩系等，后来发现奥陶系也是华北地区的一个产油层系。另外，那次预测还有一段话："太平洋外国区以内的沿岸区域可称为太平洋内带。众所周知，太平洋内带从台湾、日本、库页岛到北美加利福尼亚，为新生代沉积。边一带分布着新生代的石油矿藏。"这个结论正在验证之中。

第四章　整　　数

物质世界（至少它的一个局部）是由各种可数的结构单元组成的。例如：人类社会由一个一个的人组成，畜群由牲畜、宇宙由天体、生物由细胞、化合物由分子等组成，这些单元的数目都是自然数，所以自然数可以看作是反映客观世界本质的一种重要的秩序，也就是信息。自然数加自然数仍是自然数，因此，自然数的表达的秩序不会因加法处理而失真。减法虽然是加法的反面，但自然数减自然数未必是自然数。可见减法处理可能使自然数秩序失真，但同时却把自然数扩张为整数。整数加、减整数仍是整数。整数质表达的秩序不因加、减处理而失真。

整数集$\{x_i\}$中的元素都是数值，它们之间依一定方式相减构成差分，差分的全体构成差分系。任意个x_i互相加减，得出可公度系。差分系或可公度系表达了许多整数体系中的信息。周期性就是可公度性的一个特款。

在这一章里，我们讨论自然数系（简称N系）和整数系（简称I系）中的信息。演算方法只限于加法和减法，乘数为整数的乘法是有限多次连加或连减的简写，

它和数值互乘有概念上的区别。

§1 自然数

自然数在我们日常了解中似乎比较简单，但是严格地讲，应当用皮阿罗（Peano）公理去定义它。皮阿罗公理的要点之一是引入后继元素。用通常的语言说就是：每一自然数加1成为下一个自然数——后继元素。公理规定1不是一个后继元素，所以零不是自然数。但在可数数体系中，特别是在可数数的数组系中，零可能出现，并且没有破坏可数的秩序，因此，有时零也包含在自然数之内①。

自然数信息体系在自然科学中占十分重要的位置。近代自然科学的发展逐步揭开了自然界的结构，其中许多结构是和自然数有关的。例如原子序号，主量子数，波数，分子成官能团（如肽链）排列等。大家比较熟悉的例子是元素周期律的发现。元素周期律的主要指标是原子序号，它是自然数。早在1862年～1864年科学界就觉察到原子量具有一定可排列的规律，到了1869年门捷列夫和迈耶发表了元素周期律，并且预言镓、钪、锗等新元素的存在。镓在1875年、钪在1879年、锗在1886年

---

① Nathan Jacobson, Basic algebra 1, W. H. Freeman and Co. san Francisco p15, 1974.

陆续被发现，预测得到了证实。现在，原子序号与元素性质之间的关系已经为科学界公认。

在技术经济预测中，N 系信息也不能忽视。如预测某些小批量商品、牲畜的数量等。

§2 整　数

整数信息体系是从自然数和零的信息体系中扩张出来的。被整数计算的对象一般是数值，并且允许减法演算。和自然数一样，整数也是反映客观世界的一种重要秩序，如整数体系例一。

整数体系例一——稳定粒子质量迭加式

稳定重子和部分介子的静止质量可以近似地用 π^0、μ、$\pi^\pm$ 三种粒子的静止质量的迭加来形象。迭加公式为：

$$M(i,j,k)=i[\pi^0]+j[\mu]+k[\pi^\pm]$$

三种粒子的静止质量（单位是兆电子伏）如下

$[\pi^0]\approx134.9645$ $[\mu]=105.6595$ $[\pi^\pm]=139.5688$(Mev)

迭加结果见表4—1。

表中$\hat{M}$为实际粒子的静止质量，I 为同位旋，J 为自旋，P 为宇称。

i，j是自然数和零，k 为整数。

$i=\{0, 1, 2, \cdots, 9\}$，$j=\{1, 2, 3, 4\}$　　而且：

表4—1

迭加公式计算值					实际粒子数据				
i+k	i	j	k	M(i, j, k)	$\hat{M}$ ·(符号)	I	J	P	类别
12	0	0	12	1674.8	1672.5 (Ω^-)	0	$^3/_2$	+	重子
10	1	2	9						
8	2		6	1318.7	1321.5 (Ξ^0)	$^1/_2$	$^1/_2$	+	
	3		5	1314.1	1314.9 (Ξ^0)				
	4		4	1309.5					
	5	1	3	1199.2	1197.3 (Σ^-)	1	$^1/_2$	+	
	6		2	1194.6	1192.7 (Σ^0)				
	7		1	1190.0	1189.5 (Σ^+)				
6	8	3	−2	1117.6	1115.6 (Λ)	0	$^1/_2$	+	
4	9	4	−5	939.5	938.3 (P)	$^1/_2$	$^1/_2$	+	
					939.6 (N)				
4	2	0	2	549.1	548.8 (η)	0	0	−	介子
2	0	2	2	490.5	493.7 (K^+)	$^1/_2$	0	−	
					497.7 (K^0)	$^1/_2$	0	−	

$$\frac{\Delta k}{\Delta i}=\begin{cases} -3 & i<2,\ i>7 \\ -1 & 2\leqslant i\leqslant 7 \end{cases}$$

$$I=\begin{cases} 0 & =\{0,\ 3\} \\ \frac{1}{2} & j=\{2,\ 4\} \\ 1 & j=\{1\} \end{cases}$$

最稳定的粒子——中子N和质子P的k_p值为：

$$k_p = \mathrm{Min}(k) = -5$$

表4—1中的规律是相当明显的。如果粒子质量迭加式不是一种偶然的巧合，那末，物理学家可能会发现π^0、μ、$\pi^{\pm}$这三种粒子是组成物质世界的中间组块。相当于表4—1中i=1和i=4的粒子也可能被发现或被证明不能存在。

§3 差　　分

对于数据序列$\langle x_i \rangle = \langle x_1, x_2, \cdots, x_i, \cdots, x_n \rangle$，各阶差分是：

1阶：$\Delta x_i = x_{i+1} - x_i$

2阶：$\Delta^2 x_i = \Delta x_{i+1} - \Delta x_i = x_{i+2} - 2x_{i+1} + x_i$

……　　……

k阶：$\Delta^k x_i = \Delta^{k-1} x_{i+1} - \Delta^{k-1} x_i$

$$= x_{i+k} - \frac{k}{1!} x_{i+k-1} + \frac{k(k-1)}{2!} x_{i+k-2} + \cdots + (-1)^k x_i$$

为了形象差分之间的关系，可以用表4—2示意：

如果$\Delta^k x$最一个常数。那末，k阶以上的差分均为零。

表4—2

i	x	Δx	Δ^2x	Δ^3x	Δ^4x
i−3	x_{i-3}				
		Δx_{i-3}			
i−2	x_{i-2}		$\Delta^2 x_{i-3}$		
		Δx_{i-2}		$\Delta^3 x_{i-3}$	
i−1	x_{i-1}		$\Delta^2 x_{i-2}$		$\Delta^4 x_{i-3}$
		Δx_{i-1}		$\Delta^3 x_{i-2}$	
i	x_i		$\Delta^2 x_{i-1}$		$\Delta^4 x_{i-2}$
		Δx_i		$\Delta^3 x_{i-1}$	
i+1	x_{i+1}		$\Delta^2 x_i$		$\Delta^4 x_{i-1}$
		Δx_{i+1}		$\Delta^3 x_i$	
i+2	x_{i+2}		$\Delta^2 x_{i+1}$		
		Δx_{i+2}			
i+3	x_{i+3}				

差分的基本外推式是[①]：

$$x^*_{i+k}=x_i+\frac{k}{1!}\Delta x_i+\frac{k(k-1)}{2!}\Delta^2 x_i+\cdots+\Delta^k x_i$$

由于各阶差分的取值方法不同，可以设计出不同的预测公式。

① 伊·彼·梅索夫斯基赫（И. П. Мысовских），《计算方法》，人民教育出版社，67页，1960。

§4 可公度信息系

“可公度性”（Commensurability）一词是在天文学中首先提出来的。由于至今还没有人能够提出有说服力的机制理论，一直当作经验关系写入某些天文文献之中。可公度性是自然界的一种秩序，所以是一种信息系。很可惜，现在许多人快要把它忘记了。为了把可公度的信息系，从天文学扩张到预测科学，现在介绍一下有关史实。

太阳系几个天体的“平均运动”可公度到不平常的程度，完全超过偶然可公度的可能性。绕轨道运行天体的“平均运动”直接和它到轨道中心的主星（Primary）的距离有关。所以，“平均运动”的可公度性意味着平均距离的可公度性。最早注意到太阳系行星距离的可公度性的可能是波特（J.E.Bode，1747～1826）和提塔斯（J.B.Titius，1729～1796），波特定则可写成下列形式：

$$y_i = i \qquad i = \{(-\infty),\ 0,\ 1,\ 2,\ 3,\ \cdots\}$$

式中 i 是整数，y_i 是行星到太阳的距离 x_i（用天文单位(A.U.)计量）的函数：

$$y_i = \frac{\ln(x_i - 0.4) - \ln 0.3}{\ln 2}$$

波特定则的计算值x_i和行星的实际日星距离$\hat{x}_i$比较如表4—3：

表4—3

i	$-\infty$	0	1	2	3	4	5	6		7
x_i	0.4	0.7	1	1.6	2.8	5.2	10	19.6		38.8
$\hat{x}_i$	0.39	0.72	1	1.52	2.65	5.2	9.54	19.2	30.1	39.5
行星	水星	金星	地球	火星	小行星	木星	土星	天王星	海王星	冥王星

拉普拉斯（Pierre S.Laplace，1749～1827）注意到木星的卫星（木卫一(Io)、木卫二(Europa)和木卫三(Ganymede)）的“平均运动”x_1、x_2和x_3服从下式可公度式：

$$x_3+x_3-x_2-x_2=x_2-x_1$$

在太阳系中的其它卫星也发现有可公度性。例如土星的卫星：土卫一（Mimas）、土卫二（Enceladus）、土卫三（Tethys）和土卫四（Dione）的“平均运动”x_1，x_2，x_3和x_4服从下列可公度式（Hermann Struve）：

$$4x_4+x_3-5x_2=5x_2-5x_1$$

这一系列可公度性传递出太阳系星体公转半径的信息。

兹后，还陆续有这方面的研究报道（如：Miss M.A.Blagg，A.E.Roy，M.W.Ovenden，D.Kirkwood等）。

在天文学研究的基础上，我们提出可公度信息系的

一般表示式：

$$x_{i+1}=\sum_{j=1}^{i} I_j x_j$$

式中 $\{j\}\subseteq\{i\}$，即 j 是下标集 $\{i\}=\{1, 2, \cdots, n\}$ 中的元素，I_j 为整数。

当然，一个可公度式可能是偶然的，不能作为预测的依据。为了说明 x_{i+1} 的非偶然性，必须有一个以上的可公度式：

$$x_{(i+1),1}=\sum_{j_1=1}^{i_1} I_{j1} x_{j1}$$

$$x_{(i+1),2}=\sum_{j_2=1}^{i_2} I_{j2} x_{j2}$$

$$\cdots\cdots \qquad \cdots\cdots$$

把上述一系列 $x_{(i+1)}$ 的值排列成单调上升的半序集：

$$\langle x_{(i+1),1},\ x_{(i+1),2},\ \cdots,\ x_{(i+1),m} \rangle$$

并要求：

$$(x_{(i+1),m}-x_{(i+1),1})<\varepsilon$$

式中的 ε 是确定模型中事先确定的可行临界值。

如果满足上述要求的可公度式多于1个，即 $m>1$，那么，x_{i+1} 就可能不是偶然的。不过，为了估计 x_{i+1} 的非偶然性的程度，还要用到随机性的概念和方法。

可公度系例一——唐山地震的四元周期

唐山一带据历史记载发生过 M≥5.5 级地震6次，时

间是：$\langle \hat{x}_i \rangle = \langle \hat{x}_1 = 1527.7.1,\ \hat{x}_2 = 1568.4.25,\ \hat{x}_3 = 1624.4.17,\ \hat{x}_4 = 1795.8.5,\ \hat{x}_5 = 1805.3.12,\ \hat{x}_6 = 1945.9.23 \rangle$[1]。以12个月为一年，30日为一月换算，用可公度式求得概周期：

$$\hat{x}_4 + \hat{x}_2 - \hat{x}_5 - \hat{x}_1 = 31.2.17$$

$$\hat{x}_5 + \hat{x}_4 - \hat{x}_6 - \hat{x}_3 = 30.9.17$$

平均四元周期约为：$\bar{\Delta x} = 30$年11月27日

从$\hat{x}_6$外推一个周期，得到后一次地震时间可能是：

$$\hat{x}_6 + \Delta x = 1976.9.20$$

实际地震发生在1976.7.28　震级7.8。

可公度系例二——1982年华北干旱前的预测

1980年笔者曾预测1982年前后华北可能发生广泛干旱，干旱中心估计在山东、河南一带[2]。这一预测和以后的实际情况大体符合。当时预测的推据之一是华北、东北地带历史上5次大旱年份的五元可公度式。那五次大旱的年份是：$\langle \hat{x}_i \rangle = \langle \hat{x}_1 = 1484,\ \hat{x}_2 = 1615,\ \hat{x}_3 = 1640,\ \hat{x}_4 = 1641,\ \hat{x}_5 = 1877 \rangle$。五元可公度式是：

$$\hat{x}_5 + \hat{x}_2 + \hat{x}_2 - \hat{x}_1 - \hat{x}_3 = 1983$$

$$\hat{x}_5 + \hat{x}_5 + \hat{x}_1 - \hat{x}_3 - \hat{x}_2 = 1983$$

$$\hat{x}_5 + \hat{x}_2 + \hat{x}_2 - \hat{x}_1 - \hat{x}_4 = 1982$$

$$\hat{x}_5 + \hat{x}_5 + \hat{x}_1 - \hat{x}_4 - \hat{x}_2 = 1982$$

---

① 北京人民出版社，《地震常识》，1975。

② 翁文波，“可公度性”，地球物理学报24（2），1981，151页。

当然，实际对干旱中心的估计还要综合其它信息才能作出的。

可公度系例三——1988年中南某地可地水灾

中南某地一带常发生水灾。历史上记载了七次大水灾年份$\hat{x}_i$，见表4—4第一列（江苏地理研究所，1976），由 i＝1—7的七次大水灾年份可得三元可公度式（见表4—4第二列），表4—4中年份的可行临界值定为±1年。从可公度性定义的信息可以预测：该地1988年可能发生第8次水灾。

表4—4

x_i	三元可公度式
$\hat{x}_1=1553$	$\hat{x}_2+\hat{x}_3-\hat{x}_4=1553$，$\hat{x}_2+\hat{x}_6-\hat{x}_7=1554$，$\hat{x}_3+\hat{x}_5-\hat{x}_7=1553$
$\hat{x}_2=1566$	$\hat{x}_1+\hat{x}_4-\hat{x}_3=1566$，$\hat{x}_1+\hat{x}_7-\hat{x}_6=1565$，$\hat{x}_4+\hat{x}_5-\hat{x}_7=1566$
$\hat{x}_3=1645$	$\hat{x}_1+\hat{x}_4-\hat{x}_2=1645$，$\hat{x}_1+\hat{x}_7-\hat{x}_5=1645$，$\hat{x}_4+\hat{x}_6-\hat{x}_7=1646$
$\hat{x}_4=1658$	$\hat{x}_2+\hat{x}_3-\hat{x}_1=1658$，$\hat{x}_2+\hat{x}_7-\hat{x}_5=1658$，$\hat{x}_3+\hat{x}_7-\hat{x}_6=1657$
$\hat{x}_5=1883$	$\hat{x}_1+\hat{x}_7-\hat{x}_3=1883$，$\hat{x}_2+\hat{x}_6-\hat{x}_3=1884$，$\hat{x}_2+\hat{x}_7-\hat{x}_4=1883$
$\hat{x}_6=1963$	$\hat{x}_1+\hat{x}_7-\hat{x}_2=1962$，$\hat{x}_3+\hat{x}_5-\hat{x}_2=1962$，$\hat{x}_3+\hat{x}_7-\hat{x}_4=1962$
$\hat{x}_7=1975$	$\hat{x}_3+\hat{x}_5-\hat{x}_1=1975$，$\hat{x}_4+\hat{x}_5-\hat{x}_2=1975$，$\hat{x}_4+\hat{x}_6-\hat{x}_3=1976$ $\hat{x}_6+\hat{x}_2-\hat{x}_1=1976$
$\hat{x}_8=1988$	$\hat{x}_7+\hat{x}_2-\hat{x}_1=1988$，$\hat{x}_5+\hat{x}_4-\hat{x}_1=1988$，$\hat{x}_7+\hat{x}_4-\hat{x}_3=1988$

§5 有限整数

前文谈到了自然界许多体系、特别是物质和能量的

基本结构是由可数的单元构成，所以，整数本身就是信息。

整数也可用来传递信息，因为整数简单，容易记忆和模拟。如数列，地震烈度，风力和海浪强度，成绩考核，符号等都用整数分级或分位。

被应用的整数一般是有限的。许多问题还可以用同余类的概念将整数限制在更小的范围内。因为一个整数被另一个整数 m 除，它的余数一定是：0、1、2、…、（m －1）中的一个，因此可以把全体整数按照被 m 除的余数分为m 类。如余数为0的所有整数构成一类，它们是：0，m，2m，…，－m，－2m，…；余数为1的所有整数也构成一类，它们是：1，m＋1，2m＋1，…，－（m＋1），－（2m＋1），…，等等。这m 类整数记作 R_m〔0〕，R_m〔1〕，…，R_m〔m－1〕。称为整数模m 的剩余类。剩余类构成的集合只有m 个元素，所以称为有限域。这些剩余数也可以象普通数那样进行加法和乘法演算并使演算结果仍为被 m 除的余数。例如：m ＝7，因为3×4＝12，所以R_7〔3〕×R_7〔4〕＝R_7〔5〕。又因6＋9＝15，所以 R_7〔6〕＋R_7〔2〕＝R_7〔1〕等。

整数模2的剩余类只有R_2〔0〕和R_2〔1〕两个数，所有的奇数都是R_2〔1〕类；所有偶数都是R_2〔0〕类。有时候还可以进一步简化整数模2的剩余类，用0代表

$R_2[0]$，用1代表$R_2[1]$，那么，向量（3,3,4）=（1，1，0）。

用剩余类传递信息的技术，现在还只是在线性码的编码问题等方面，如校验矩阵的演算上得到了应用。

第五章　预知信号

电报译码要用预知的译码本，无线电的调频或调幅传讯也要用预知的解调信号。在海潮预测中，起潮的主要机制已确知是日、月、地之间的力学关系。作为海潮预测主要依据的这种力学关系也属于预知信号的范畴。

在有些信息体系中，我们能预知的信号可能不那么明显，有时只知道一个大概，如：一个频率范围，一个波形，甚至只知道某种信号的存在。

自相关和自褶积分析是检查数据序列中可能存在某种秩序（或信号）的方法之一。如果已经预知信号的解调序列，还常常会用到相关、褶积、反褶积等分析方法来提取信号。这类方法都要用到乘法演算。不过，这里的乘法与第四章中谈到的乘法不同，它不是连加或连减的简写，而是数值对数值的互乘。

§1 乘法和乘法表

在一定域内，乘法和加法有同构关系，如数值的乘

法可以变换成对数的加法，但在具体的信息处理中，这两种演算的作用并不相同。例如：$\cos(t_1+t_2)$是有意义的，但$\cos(t_1 \cdot t_2)$就是未必有普遍意义的了。因此，不同的运算会影响信息传递中的效果。

为了表达非负效（例如绝对误差就是非负的），加法可取绝对值，乘法可用偶次方。在计算中，乘方运算有时比绝对值运算方便。

设有一一对应的两个序列〈a，-a〉和〈b，-b〉，对应项的积是〈ab，ab〉，其平均值也是ab，而对应项的和是〈a+b，-(a+b)〉，其平均值就是零了，也就是说，要表达对应序列的相关程度，乘法可以用负负得正的原则，而加法就要取绝对值才行。

表5—1

	x_1	x_2	……	x_i	……	x_n
x_1	x_1^2	$x_2 \cdot x_1$	……	$x_i \cdot x_1$	……	$x_n \cdot x_1$
x_2	$x_1 \cdot x_2$	x_2^2	……	$x_i \cdot x_2$	……	$x_n \cdot x_2$
…	…	…	……		……	……
x_i	$x_1 \cdot x_i$	$x_2 \cdot x_i$	……	x_i^2	……	$x_n \cdot x_i$
…	…	…	……	…	……	…
x_n	$x_1 \cdot x_n$	$x_2 \cdot x_n$	……	$x_i \cdot x_n$	……	x_n^2

表5—1给出了有限序列〈x_i〉=〈x_1，x_2，…，x_i，…，x_n〉的乘法表，如果序列是自然数：（1，2，3，…，9），那末，表5—1就是九九表。

表中x_1^2，x_2^2，⋯，x_i^2，⋯，x_n^2这一斜行称为主对角线。平行于主对角线的元素的平均值称为τ级自相关函数值：

$$r_\tau = \frac{1}{n-\tau}\sum_{i=1}^{n-\tau} x_i \cdot x_{\tau+i} \qquad \tau = 0, 1, \cdots, (n-1)$$

$\langle r_0, r_1, \cdots, r_{n-1} \rangle$ 称为自相关序列。

τ级二阶矩自相关系数是：

$$\rho_\tau = \frac{\sum_{i=1}^{n-\tau} x_i \cdot x_{\tau+i}}{\sqrt{\left(\sum_{i=1}^{n-\tau} x_i^2\right)\left(\sum_{i=1}^{n-\tau} x_{\tau+i}^2\right)}}$$

垂直于主对角线的各元素的平均值称为 c 级自褶积函数值：

$$C_c = \frac{1}{c-1}\sum_{i=1}^{C-1}(x_i \cdot x_{c-i}) \qquad c = 2, 3, \cdots, (n+1)$$

$\langle C_2, C_3, \cdots, C_{n+1} \rangle$ 称为自褶积序列。

与τ级二阶矩自相关系数对应的还可以定义c级自褶积系数：

$$\mathfrak{C}_c = \frac{\sum_{i=1}^{C-1}(x_i \cdot x_{C-i})}{\sqrt{\left(\sum_{i=1}^{C-1} x_i^2\right)\left(\sum_{i=1}^{C-1} x_{C-i}^2\right)}}$$

自相关、自褶积例一——北京旱涝周期

中国许多地区500年来的历史气象记录已整理为五

级分级体系，5级表示干旱，4级表示偏旱，3级中常，2级偏涝，1级表示涝。对这种五级分级的时间（年度）序列，试作自相关和自褶积分析。

作为对照，先讨论：1）均匀随机序列，代表无信息序列。2）纯信息序列，代表无干扰序列。

1）均匀随机序列。假设一个均匀随机序列：$\langle x_i \rangle = \langle 2, 5, 3, 1, 4, 1, 3, 4, 2, 5, 3, 4, 1, 5, 2, 4, 2, 5, 3, 1 \rangle$。相应的自相关序列 r_τ 和自褶积序列 C_c 见图 5—1。图中实线表示自褶积 C_c，虚线表示自相关 r_τ。

数集 $\langle x_i \rangle = \langle x_j \rangle = \langle 1, 2, 3, 4, 5 \rangle$ 自乘积的平均值是：

$$\frac{1}{25}\sum_{i=1}^{5}\sum_{j=1}^{5} x_j x_i = 9$$

图5—1 中横线表示 $r_\tau = \bar{C}_c = 9$，整个 r_τ 和 C_c 序列在横线上下均匀分布。

2）纯信息序列。假设一个纯信息序列：$\langle x_i \rangle = \langle 1, 2, 3, 4, 5, 1, 2, 3, 4, 5, 1, 2, 3, 4, 5, 1, 2, 3, 4, 5 \rangle$。相应的自相关序列 r_τ 和自褶积序列 C_c 见图 5—2，其图例同图5—1。

图5—2中，自相关序列 r_τ 具有周期 $\Delta\tau = 5$ 的3个峰，$\tau = \{5, 10, 15\}$，3个谷，$\tau = \{3, 8, 13\}$，全序列在

$r_\tau=9$ 的横线上下分布。自褶积序列C_c 则有周期 $\Delta c=5$ 的 3 个峰，c={8，13，18}，2 个谷，c={11，16}，全序列几乎都在 $C_c=9$的横线之下。

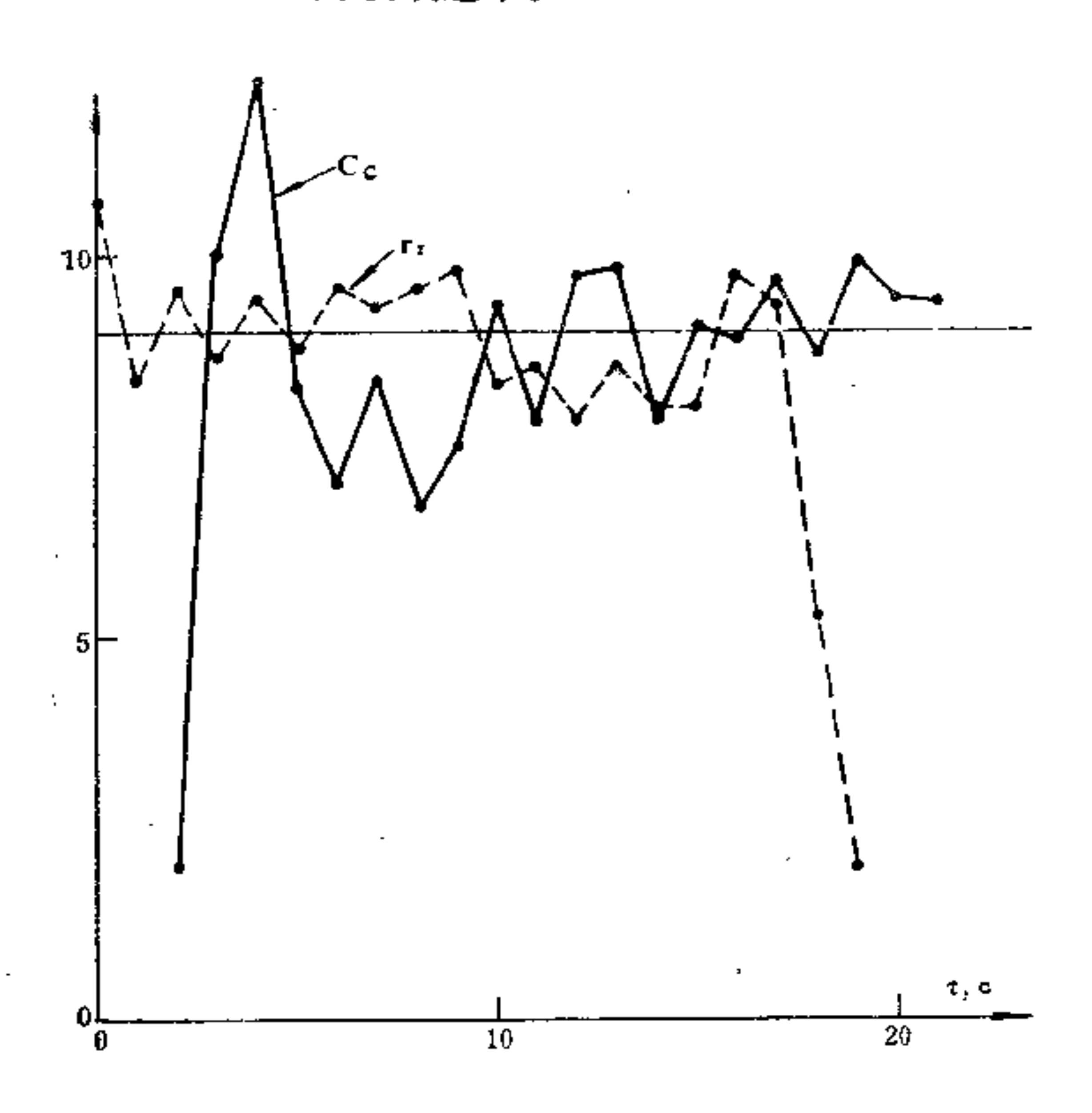


图5—1　均匀随机序列自相关、自褶积序列

3）实际序列：北京从1955～1974年20年间的旱涝五级分级序列是：

$\langle x_t \rangle = \langle 2, 1, 4, 3, 1, 4, 3, 4, 2, 3, 5, 3, 3, 4, 2, 3, 4, 4, 2, 3 \rangle$。相应的自相关序列 r_τ 和自褶积序列 C_c 见图 5—3，共图例同图 5—1。

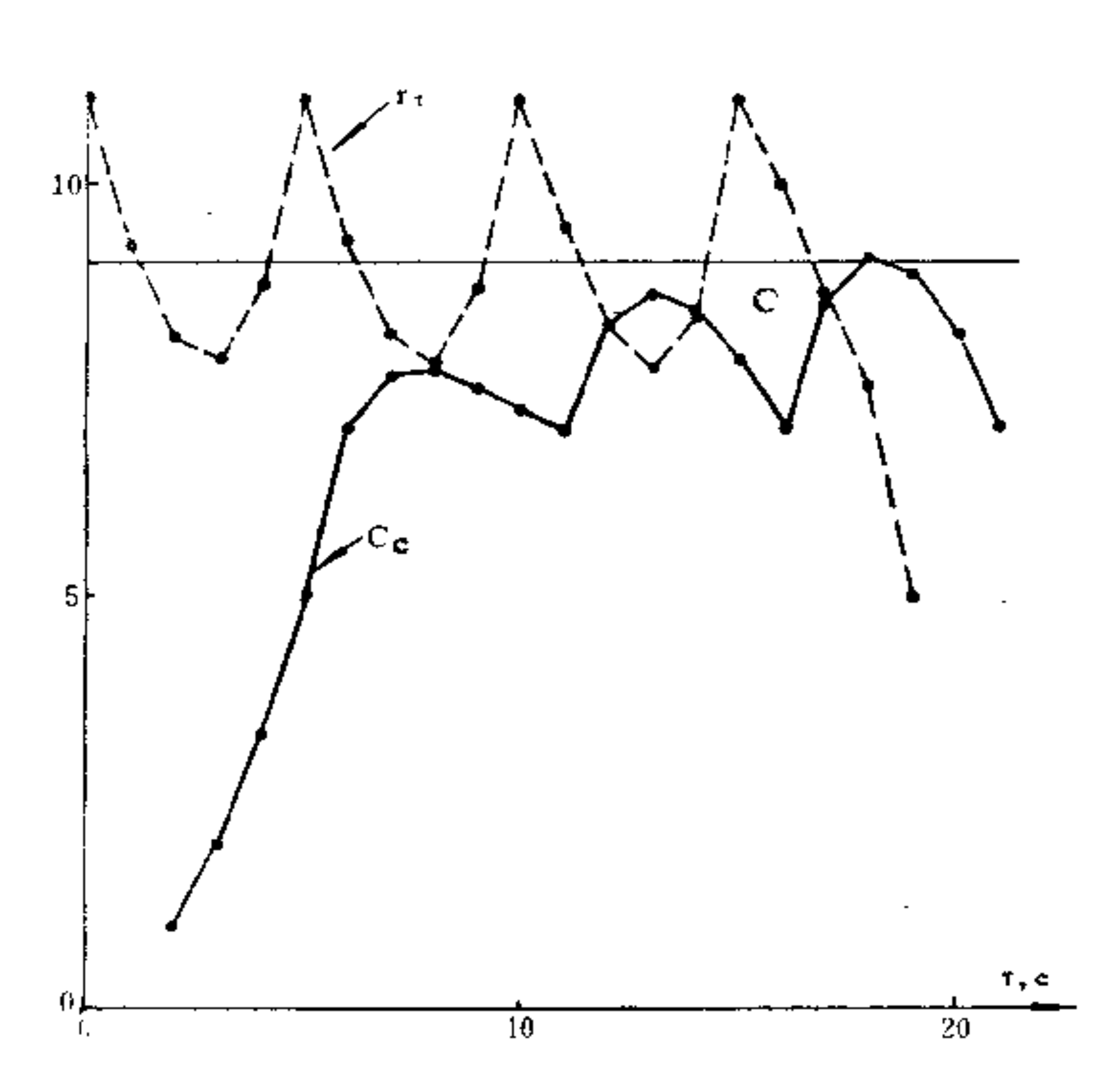


图5—2　纯信号序列自相关、自褶积序列

图5—3中，自褶积序列 C_c 大部位于 $C_c = 9$ 的横线以下。显示出很微弱的 $\Delta c = 5$ 的周期性。谷位于 $c = \{3, 8, 13, 18\}$，峰位于 $c = \{4, 9, 14\}$。本例只取了 500

年史历记载中的20年的资料，因此结果只是示意性的，但可说明，从全部资料中有可能取得周期性的信息。

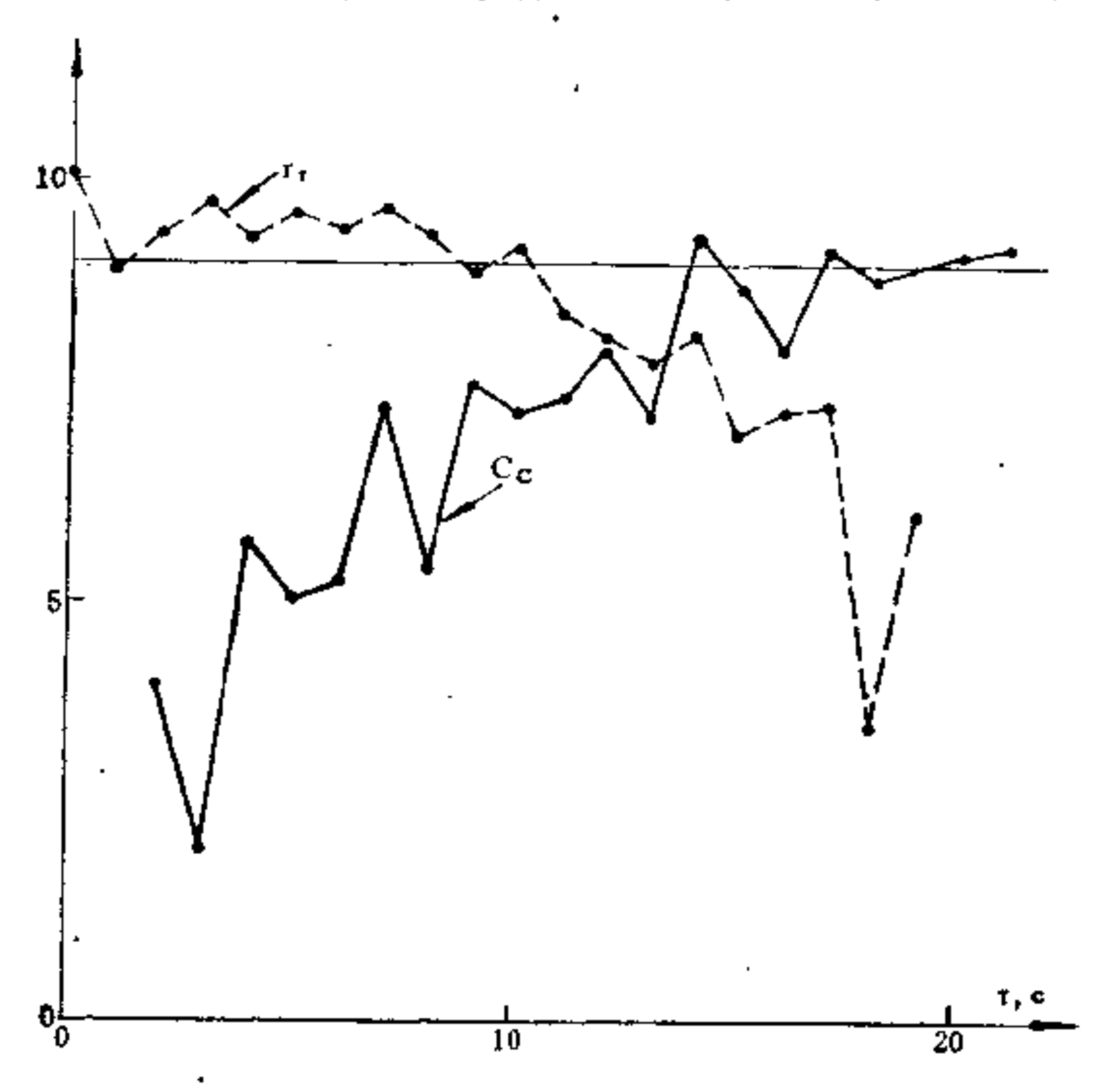


图5—3　北京地区气候分级时间序列自相关、自褶积序列

§2 周期函数多项式

周期性是数据序列中的一种重要秩序。一个没有

周期性信号的零和平稳序列，它们的自相关系数ρ_τ随着τ的增加逐步下降。反之，含有显著周期性信号的零和序列，其自相关系数ρ_τ和自褶积系数$\tilde{C}_c$的图上都会表显出峰和谷。如果周期性因素很复杂，许多峰和谷互相混淆，显不出来，但自相关系数ρ_τ和自褶积系数C_c一般不随τ和c的增加逐步下降。

为了分析含有周期性序列的秩序，通常用周期性函数多项式作模型。多项式的每一项都是一个周期函数，其中包括三个主参量。这三个主参量可以是：振幅参量，频率参量和初相位参量。也可以是余弦振幅，正弦振幅和频率。要同时确定各项的三个参量是很困难的，一般都要对其中一个参量作某种假设，然后再定量估计其它两个参量。在傅里叶级数展开式中，有两个有关频率的规定：（1）规定频率是谐和的，也就是说每一项的频率都是某个基频的自然数的倍数。（2）基频就是序列全域宽度的倒数。

设有一个零和，等节点距离的序列：

$$\langle x_i \rangle = \langle x_1, x_2, x_3, \cdots, x_i, \cdots, x_n \rangle$$

相应的半幅傅里叶级数模型可以写成：

$$x_i = \sum_{j=1}^{n/2} a_j \cos\left(\frac{2\pi j}{n} i + C_j\right) \qquad (\text{n 为偶数})$$

因为$\langle x_i \rangle$是等节点距离的，i可以代表时间，那么，频率就是：

$$\langle f_i \rangle = \langle \frac{j}{n} \rangle = \langle \frac{1}{n}, \frac{2}{n}, \cdots, \frac{j}{n}, \cdots, \frac{1}{2} \rangle$$

如果序列中的特性频率不接近上述离散的谐和频率中的任一个，那么这样的模型就会使频率信息严重失真。

为了减少这种频率信息失真的程度，建议采用“浮动频率”提取信号的方法[①]。

§3 互　乘　表

从乘法表可以扩张为两个数字序列 $\langle d_i \rangle$ 和 $\langle g_j \rangle$ 的互乘表（见表5—2）。

表5—2

	d_1	d_2	d_3	…	d_m	…	d_i	…	d_n
g_1	d_1g_1	d_2g_1	d_3g_1	…	d_mg_1	…	d_ig_1	…	d_ng_1
g_2	d_1g_2	d_2g_2	d_3g_2	…	d_mg_2	…	d_ig_2	…	d_ng_2
…	…	…	…	…	…	…	…	…	…
g_j	d_1g_j	d_2g_j	d_3g_j	…	d_mg_j	…	d_ig_j	…	d_ng_j
…	…	…	…	…	…	…	…	…	…
g_m	d_1g_m	d_2g_m	d_3g_m	…	d_mg_m	…	d_ig_m	…	d_ng_m

设 $\langle d_i \rangle = \langle d_1, d_2, \cdots, d_i, \cdots d_n \rangle$

① 翁文波，“频率信息的保真”，石油地球物理勘探，1980“3”1页。

$\langle g_j \rangle = \langle g_1, g_2, \cdots, g_j, \cdots, d_m \rangle \qquad (n > m)$

和乘法表相似，τ 级互相关函数值是：

$$r_\tau = \begin{cases} \dfrac{1}{m}\sum_{j=1}^{m} d_{j+\tau} \cdot g_j & 0 \leqslant \tau \leqslant (n-m) \\ \dfrac{1}{n-\tau}\sum_{j=1}^{n-\tau} d_{j+\tau} \cdot g_j & (n-m) \leqslant \tau \leqslant (n-1) \end{cases}$$

互相关序列是：

$$\langle r_0, r_1, \cdots, r_{(n-1)} \rangle$$

如果$\langle d_i \rangle$和$\langle g_j \rangle$一一对应，即$n=m$，那么$\tau=0$时的两个序列的二阶矩相关系数是：

$$\rho_{d,g} = \frac{\sum_{j=1}^{m} d_j g_j}{\sqrt{\left(\sum_{j=1}^{m}(d_j)^2\right)\left(\sum_{j=1}^{m}(g_j)^2\right)}}$$

同样，C 级互褶积函数值为：

$$C_C = \begin{cases} \dfrac{1}{c-1}\sum_{j=1}^{C-1} d_{C-j} \cdot g_j & 2 \leqslant c \leqslant (m+1) \\ \dfrac{1}{m}\sum_{j=1}^{m} d_{C-j} \cdot g_j & (m+1) \leqslant c \leqslant (n+1) \end{cases}$$

褶积序列为：

$$\langle C_2, C_3, \cdots, C_{(n+1)} \rangle$$

举一个预知脉冲信号的例子。设预知原来的脉冲信号序列为：〈1，0，0，…〉。又根据理论或实际资料的演算总结等方法，取得了标准信号序列为：$\langle 1, x_2,$

$x_3, \cdots$ 〉。现在来求褶积的解码序列，使它与标准信号序列的褶积为脉冲信号序列。

引入变量Z和解码序列〈 $D_1, D_2, \cdots$ 〉使：

$$D_1+D_2Z+D_3Z^2+D_4Z^3=1/\left(1+x_2Z+x_3Z^2+x_4Z^3+\cdots\right)$$

令Z的等幂项系数相等得到：

$D_1=1$　　　　$D_2=-x_2$

$D_3=x_2{}^2-x_3$　　　　$D_4=-(x_2{}^2-2x_2x_3+x_4)$

····　　　　····

作互乘表（表5—3）

表5—3

	1	x_2	x_3	x_4	···
1	1	x_2	x_3	x_4	···
$-x_2$	$-x_2$	$-x_2^2$	$-x_2x_3$	···	···
$x_2^2-x_3$	$x_2^2-x_3$	$x_2^3-x_2x_3$	···	···	···
$-(x_2^3-2x_2x_3+x_4)$	$-x_2^3+2x_2x_3-x_4$	···	···	···	···
···	···	···	···	···	···

表5—3的褶积序列就是〈1，0，0，···〉

由此可知〈 $D_1, D_2, D_3\cdots$ 〉就是所求的解码序列，也有人称之为反褶积因子。

§4 广义定标预测

如果只有信息的定义是相对明确的，譬如：许多单位、企业或个人所争取或追求的目标——超额完成任务，争取最高利润和最高经济效益，争夺并占据市场等——都属于信息，而其余的都是噪音。但是产生这类信息的体系却不完全清楚，那就产生广义定标预测问题。在许多预测程序中都包括广义定标预测技术的组成部分。例如下列各类技术：

（1）德尔菲技术：这是一种反复向专家个别调查的技术。包括选定专家，确定要调查的问题，明确预测目标（即明确信息的定义），使专家们保持兴趣和合作态度的办法，成果利用等。在全过程中，要估计情况，反复征求意见，公开初步结果并征求修改意见等。

（2）定标技术：狭义的定标技术一般用于较具体的问题（如预测将来可能制造成功的某种产品）。在这类问题中，事先要估计社会需要、生产技术、产品性能等各种情况。估计的方法可以向专家征求意见，或作市场调查，或建立经济模型等。在定标技术中，常常用外貌分析、关系分析等方法。外貌分析可以形象为一个表格。纵向分列各个制作阶段或工序，横向分列一切可能选择的方案，然后用一条折线将各工序的最优方案连接

起来作为整个工序的设计纲领。关系分析可以形象为关系树。用树形图列出可能涉及的各个部门及其分枝对于定标事物的关系，并主观预定或预测关系系数。关系树有助于决策中估计主要和次要的对象。

（3）普查技术：对于那些可以把全部注意力聚焦到一个很具体的信息问题时，可以用普查技术。所用的方法有：扫描、监视、追踪等。

第六章　拟合信号

前一章所说的预知信号表明，信号的物理意义和定义至少有一部分是已确定的或者可以推理得到的。那样看待信息可以说是偏于唯理的。这一章将从另一个方向出发。即对于信息的定义和性质不作任何事先的假设，而是从实际情况——现在指的是数据序列——里面找出信息来，这样看待信息可以说是偏于唯象的。

拟合就是建立一个模型去逼近实际数据序列的过程。本文只考虑确定模型，文中引入的正态分布密度函数和泊松分布概率函数也是当作确定性函数来考虑的，和概率论无关。这是名词的借用。

事物发展的兴起或盛衰过程称为生命旋回。许多生命总量不限的体系，在兴起阶段，可形象为正态旋回；另一些生命总量有限的体系，其盛衰的全过程可以形象为泊松旋回；对于有极限的体系，在邻近极限的阶段，可以形象为逻辑斯谛旋回。

从唯象的拟合信息中，只能推测到取得最后数据（信息）以前，信源的状态。以这种拟合信息为基础的预测称为“基值预测”。对于许多体系，特别是有关人

类活动的体系，基值预测和未来的实践可以有区别，因为人类永远不会满足于历史，并不断创造出前所未有的事物。所以凡对人类有利的事物体系，基值预测将会落后于实践。

§1 拟合

一个预测模型的建立要尽可能符合实际体系，这个原则可称为拟合原则。符合的程度可以有多种标准，例如：最小二乘方、最大似然性、最小绝对偏差（MAD）等。建立模型时，要求目标函数满足某种标准的最大或最小等最优化条件。由于标准和计算方法不同，结果是多种多样的。当然，建立这类模型也不是完全没有其它取舍原则，例如，历史经验可以是一种重要的依据。

拟合常用于发展性的体系，例如：新事物的兴起，或从兴起到顶峰（如生命过程的典型曲线），或从兴起到衰亡等。最常见的是多次方多项式模型。

设有数据序列：

$$\langle x_i(t_i)\rangle=\langle x_1(t_1),\ x_2(t_2),\ \cdots,\ x_i(t_i),\ \cdots,\ x_n(t_n)\rangle$$

一元 l 次多项式模型是：

$$x_i=a_0+a_1t_i+a_2t_i^2+\cdots+a_jt_i^j+\cdots+a_lt_i^l$$

估计 $(l+1)$ 个系数 a_0，a_1，a_2，…，a_j，… a_l 的矩

阵方程是：

$$
\begin{pmatrix}
n & \Sigma t_i & \cdots & \Sigma t_i^j & \cdots & \Sigma t_i^l \\
\Sigma t_i & \Sigma t_i^2 & \cdots & \Sigma t_i^{j+1} & \cdots & \Sigma t_i^{l+1} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\Sigma t_i^j & \Sigma t_i^{j+1} & \cdots & \Sigma t_i^{2j} & \cdots & \Sigma t_i^{l+1} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\Sigma t_i^l & \Sigma t_i^{l+1} & \cdots & \Sigma t_i^{j+l} & \cdots & \Sigma t_i^{2l}
\end{pmatrix}
\begin{pmatrix} a_0 \\ a_1 \\ \cdots \\ a_j \\ \cdots \\ a_l \end{pmatrix}
=
\begin{pmatrix} \Sigma x_i \\ \Sigma t_i x_i \\ \cdots \\ \Sigma t_i^j x_i \\ \cdots \\ \Sigma t_i^l x_i \end{pmatrix}
$$

式中：“Σ”是“$\sum_{i=1}^{n}$”的简写。

在经济预测中[1]，常用到 $l=2$ 的二次方模型。

此外，还可以有许多特殊形式的拟合模型，例如，时间序列 $x(t)$ 的模型可以是：

1） $x=x_0(1+\alpha)^t$

2） $x=a+bT_1(t)+cT_2(t)$

3） $\ln x=a+b\ln t$

4） $\ln x=a+b/t$

5） $\ln x^n=a+bt^m$ （n，m，常取1—5的数）

6） $\ln x=a+be^t$

7） $x=1/(a+be^t)$

8） $x=a(1-be^{Ct})$

……

① 陈善林，《经济预测法》，上海人民出版社，77页，1982。

§2 生命旋回

在预测技术中，生命旋回常称为兴衰周期、生命周期等。生命旋回是事物从兴起（如商品投入市场）、经过成长、成熟到衰退的全过程。例如：商品生命直接受需要（销售量）和利润（获利能力）的控制，又间接受价格、社会经济、科技水平、市场竞争、供需平衡等互相依存的许多因素的影响。多种因素的相消相成使商品生命显示出某种共性。

许多生命旋回体系的前一阶段，即从兴起到成熟的“S”曲线阶段，可以被形象为正态分布密度函数[①]：

$$x_t = ae^{-\frac{1}{2}\left(\frac{t-t_0}{\sigma}\right)^2}$$

其中 a 是体系参量，t_0 是x_t 极大时的 t 值，σ^2是相当于“方差”的一个参量。这类生命旋回本文称为正态旋回。正态旋回常可拟合生命总量不受直接限制的体系在达到满和前一阶段的形象。

对于生命总量有限的许多体系，如非再生资源，全生命过程可以形象为泊松分布概率函数：

$$\frac{x_t}{\sum_\infty x_t} = \frac{t^n e^{-t}}{n!}$$

① 郭军元，《市场学》，机械工业出版社，55页，1981。

其中$\Sigma_{\infty}x_{i}$为发展总量，n为常量（暂假设为整数）。这类生命旋回本文称为泊松旋回。

不同体系在生命旋回邻近极限阶段时，可以形象为逻辑斯谛（Logistic）函数：

$x_t = A/(1 + ae^{bt})$　　　　a，b 为常数

本文称之为逻辑斯谛旋回。

§3 正态旋回

正态旋回适用于生命旋回的发展阶段，或一部分衰退阶段。正态旋回以下列假设为依据：

1）正态旋回体系有一个饱和点，在饱和点之前，体系是发展的，代表值Q增加；在饱和点之后，体系是衰退的，代表值Q减少。如果把时间横坐标轴t的原点定在饱和点上，那么当$t<0$时，$\frac{dQ}{dt}>0$，当 $t>0$时 $\frac{dQ}{dt}<0$。同时假定，Q的变化速度和时间t到原点的时间间隔成正比，也就是说，离饱和点越近Q的变化越小。

2）体系Q的发展速度$\frac{dQ}{dt}$和体系Q本身成正比，也就是说基数越大，体系随时间的递增数也越大。这一假设可以在客观世界中直接观察到，例如：树林越大，每年增加的本材越多。

基于上述两点假设，可以列出以下微分方程：

$$\frac{dQ}{dt}=-Qt$$

解上列方程，得到

$$\ln Q=-\frac{t^2}{2}+\ln Q_0$$

或者： $Q=Q_0e^{-t^2/2}$

上式与正态密度函数形式完全相同。

如果令 $t=(y-y_0)/\sigma$ 则上式可以写为：

$$Q=A\cdot f\left(\frac{y-y_0}{\sigma}\right)=Af(t)$$

式中$f(t)$是标准正态密度函数。y是时间，如年份。y_0是饱和点的时间。σ是相当于“方差根”的一个时间单位变换参量。

上式有下列特点：

1）因为正态旋回一般是用于事物发展的阶段，即所谓“S”阶段，所以，通常情况下：

$t\leqslant 0$，即$y\leqslant y_0$

当$t=0$，即$y=y_0$时，

$$Q=Q_{Max}\cong 0.3989A$$

也就是说：y_0年，Q 达到饱和值。

2）$f(t)$ 对 t 的二次导数。

因为：$f(t)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2}$

所以：

$$\frac{d^2f(t)}{dt^2}=(t^2-1)f(t)=\begin{cases}>0 & t<-1\\ 0 & t=-1\\ <0 & 0>t>-1\end{cases}$$

即：Q的加速增加或猛升阶段是：

$$\frac{d^2f(t)}{dt^2}>0,\ t<-1,\ y<(y_0-\sigma)$$

Q的减速增加或一般上升阶段是：

$$\frac{d^2f(t)}{dt^2}<0,\ t>-1,\ y_0>y>(y_0-\sigma)$$

3）如果已知Q_i对y_i的三套数据，如：

$\{(Q_1, y_1), (Q_2, y_2), (Q_3, y_3)\}$ 并有

$$y_1-y_2=y_2-y_3>0$$

就可以估计有关参量：

$$y_0=\frac{y_1+(1-a)y_2-ay_3}{2(1-a)}$$

其中：

$$a=\frac{\ln\left(\frac{Q_1}{Q_2}\right)}{\ln\left(\frac{Q_2}{Q_3}\right)}$$

并有：

$$\sigma^2=\frac{(y_1-y_2)(4y_0-y_1-2y_2-y_3)}{2\ln\left(\frac{Q_1}{Q_3}\right)}$$

正态旋回例一——全国灯泡生产

中国灯泡年生产量Q① 可用正态密度函数拟合：

$$Q=38f(t)$$

$$f(t)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2}$$

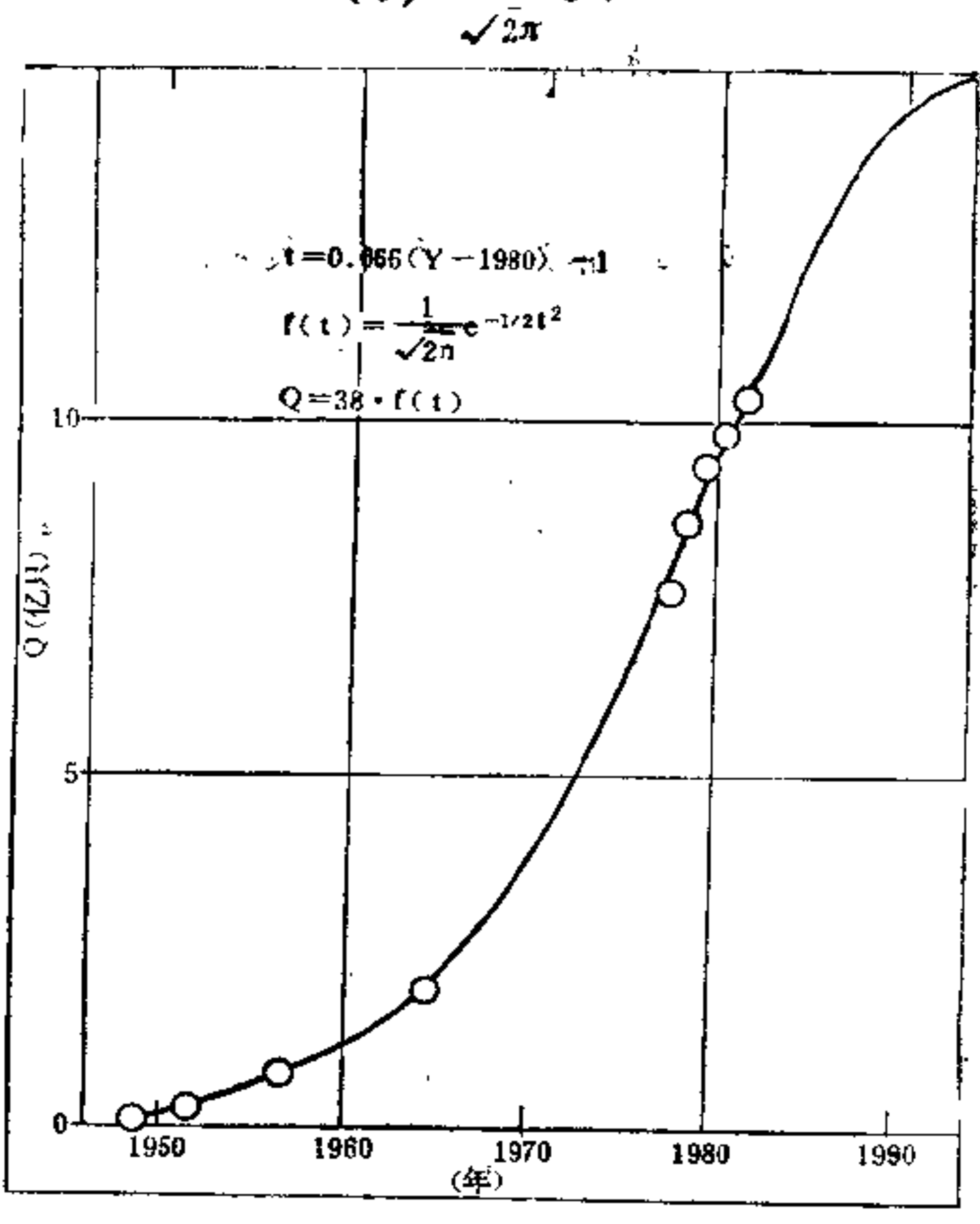


图6—1　中国灯泡生产量

① 国家统计局，《中国统计摘要》.中国统计出版社，41页，1983。

$$t = 0.066(y - 1980) - 1$$

式中y是公元年份。

计算的Q值曲线和实际值（用圆圈表示）见图6—1。从图6—1可见：如不更新换代，1995年左右灯泡的年产量将达到15.2亿只的饱和水平，这一预测也要求灯泡生产进行较基本性的技术革新。

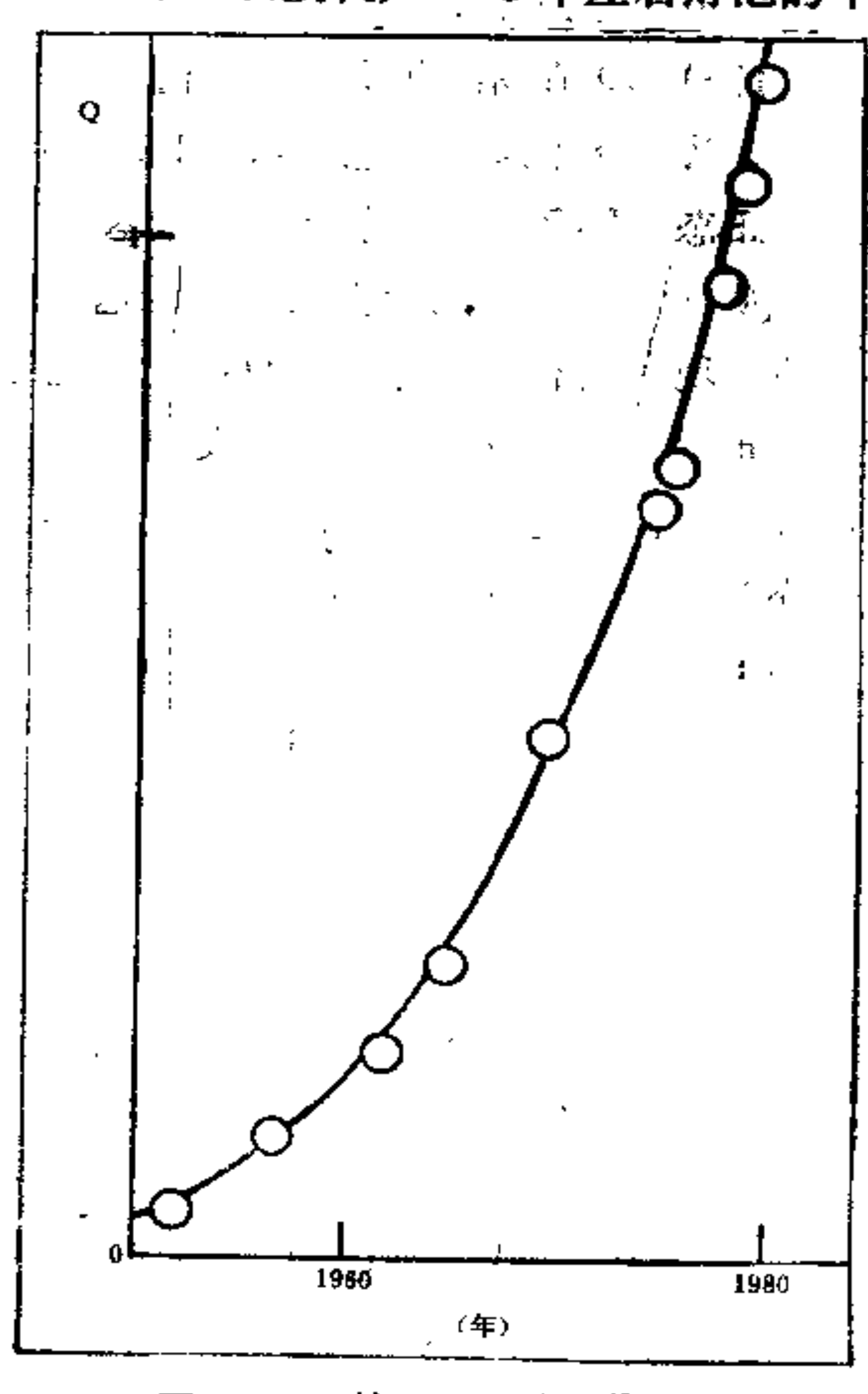


图 6—2 某市工业总产值

正态旋回例二——某城市工业总产值

某市工业总产值Q也可以用正态密度函数拟合：

$$Q = 0.032 + 4.62f(t)$$

$$f(t)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2}$$

$$t=(y-1996)/16$$

式中 y 是公元年份。

计算的Q 值曲线和实际的圆圈符号见图6—2。

Q以1978年的工业总产值为1个单位。

正态旋回例三——世界人口数基值

从1978年以来，世界人口问题越来越受到全世界的关注。现在已有不少专题刊物研究讨论这一问题。

世界人口数（记作Q）的发展，大体可分为几个阶段。公元前可能自成阶段。从公元零年到工业革命（暂以1800年为界）约1800年时间，可划为一个阶段。这一阶段的人口基值（在当时的社会和生活条件下的估计值）可用正态密度函数拟合：

$$Q=30\ f(t) \qquad (亿人)$$

$$f(t)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2}$$

$$t=(y-3380)/1580 \qquad 0\leqslant y<1800$$

式中 y 是公元年份。计算值Q 和从通用文献① 上查得的实际数值 $\hat{Q}$ 比较如下（见表6—1）：

从表6—1可以看出：工业革命发生后，实际人口

① Encyclopedia America, 122, p410, 1980.
Encyclopedia Britannica, 14, p165, 1974.

表6—1

y（公元年份）	0	1000	1600	1650	1750	1800
Q（亿人）	1.5	3.9	6.4	6.6	7.0	7.3
$\hat{Q}$（亿人）	1.5	3.4	5	5.46	6.9	9

$\hat{Q}\cong 9$亿人，已明显超过计算值$Q\cong 7.3$亿人。从以上模型外推，也就是假设工业革命没有发生，世界继继处于中世纪状态，到3380年人口才达到12亿人的顶峰。

工业革命以后，约从1800年到1980年的180年，又可以划为一个阶段，这一阶段人口基值的正态旋回拟合公式是：

$$Q\cong 9.3+134f(t) \qquad \text{（亿人）}$$

$$f(t)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2}$$

$$t=(y-2045)/65 \qquad 1800\leqslant y<1980$$

上式的计算值Q和外推趋势见图6—3中的曲线。圆圈是实际值$\hat{Q}$。如果世界继续处于19～20世纪的状态，世界人口基值将于2045年达到63亿人的顶峰。外推到2000年，世界人口基值$Q\cong 51$亿可能会被实际人数超过。马勒[1]和毛尔丁[2]等人估计，实际值将在60亿左右。

① Holfdan. Mahler: Scientic American 243(3), 1980 (Sept.).

② W. Parker, Mauldin: Science 209, 1980 (No V.).

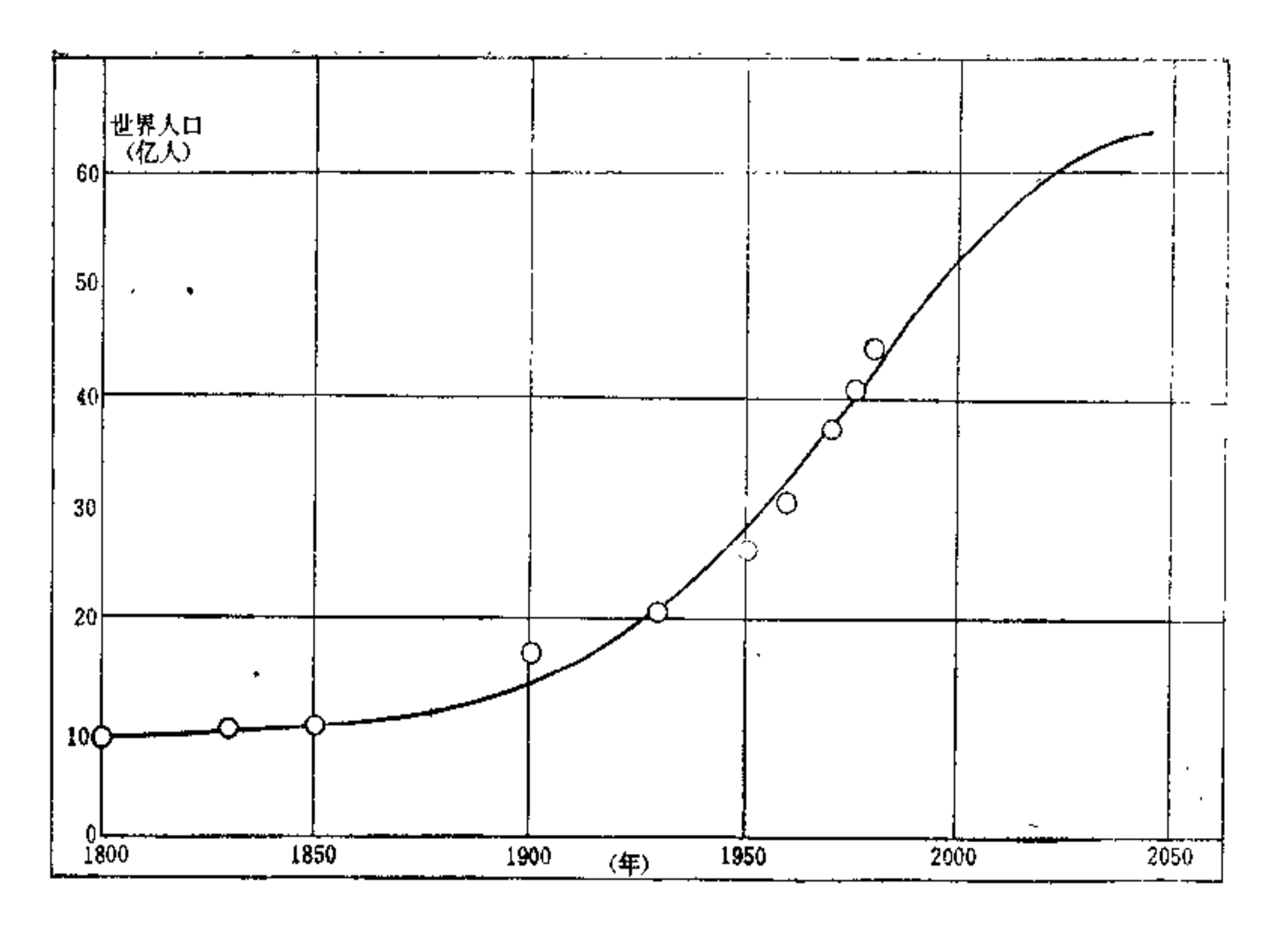


图6—3　世界人口数基值

§4 泊松旋回

假设一件事物Q在随时间t的变化过程中，正比于t的n次方函数兴起，又随着t的负指数函数衰减，这种过程可以用下列函数表示：

$$Q_t = At^n e^{-t} \qquad t \geqslant 0$$

这一函数具有下列性质：

1） $$\frac{dQ_t}{dt} = Q_t\left(\frac{n}{t} - 1\right)$$

当$t > n$时　$dQ_t/dt > 0$

$t = n$时　$dQ_t/dt = 0$

$t < n$时　$dQ_t/dt < 0$

2） $$\frac{d^2Q_t}{dt^2} = Q_t \cdot \frac{1}{t^2}\left[(t-n)^2 - n\right]$$

当$t = n \pm \sqrt{n}$　时　$d^2Q/dt^2 = 0$

3） $$\int_0^\infty Q_t dt = A \cdot n! = \Sigma_\infty Q_t$$

4） $$Q_t/\Sigma_\infty Q_t = \frac{t^n e^{-t}}{n!}$$

这就是单项泊松分布概率函数。积分式见附录。

从以上性质可知，事物Q的盛衰可分成四个阶段：

1）加速上升阶段：　$t = 0$ —— $(n - \sqrt{n})$

2)一般上升阶段： $t=(n-\sqrt{n})$ —— n

3)一段下降阶段： $t=n$ —— $(n+\sqrt{n})$

4)缓慢下降阶段： $t=(n+\sqrt{n})$ —— ∞

如$t=t_1$ 和$t=t_2$，分别有Q_{t1} 和Q_{t2} ，则：

$$n=\frac{(t_2-t_1)+\ln\frac{Q_{t2}}{Q_{t2}}}{\ln\left(\frac{t_2}{t_1}\right)}$$

这个模型是收敛的，只能用于有限体系，如矿产资源等。

泊松旋回例——世界石油产量基值的宏观预测

世界石油年产量Q_y受到地下资源、世界经济、探采技术等多种因素的控制，并且受到国际政治、军事等因素的强烈影响。地下可采储量、探采技术、和其它能源的开发也在不断发展着。因此作超远程预测是十分困难的。下面只就某些条件固定不变的假设下作基值预测。

在最终可采储量和采收率面定的假设下，一个世界石油年产量基值（Q_y）的模型如下：

$Q_y=ab^te^{-t}$ （亿吨/年）

$t=(y-1918)/10$

$a=0.04$

$b=7$

式中y 为公元年份，$y\geq 1918$。

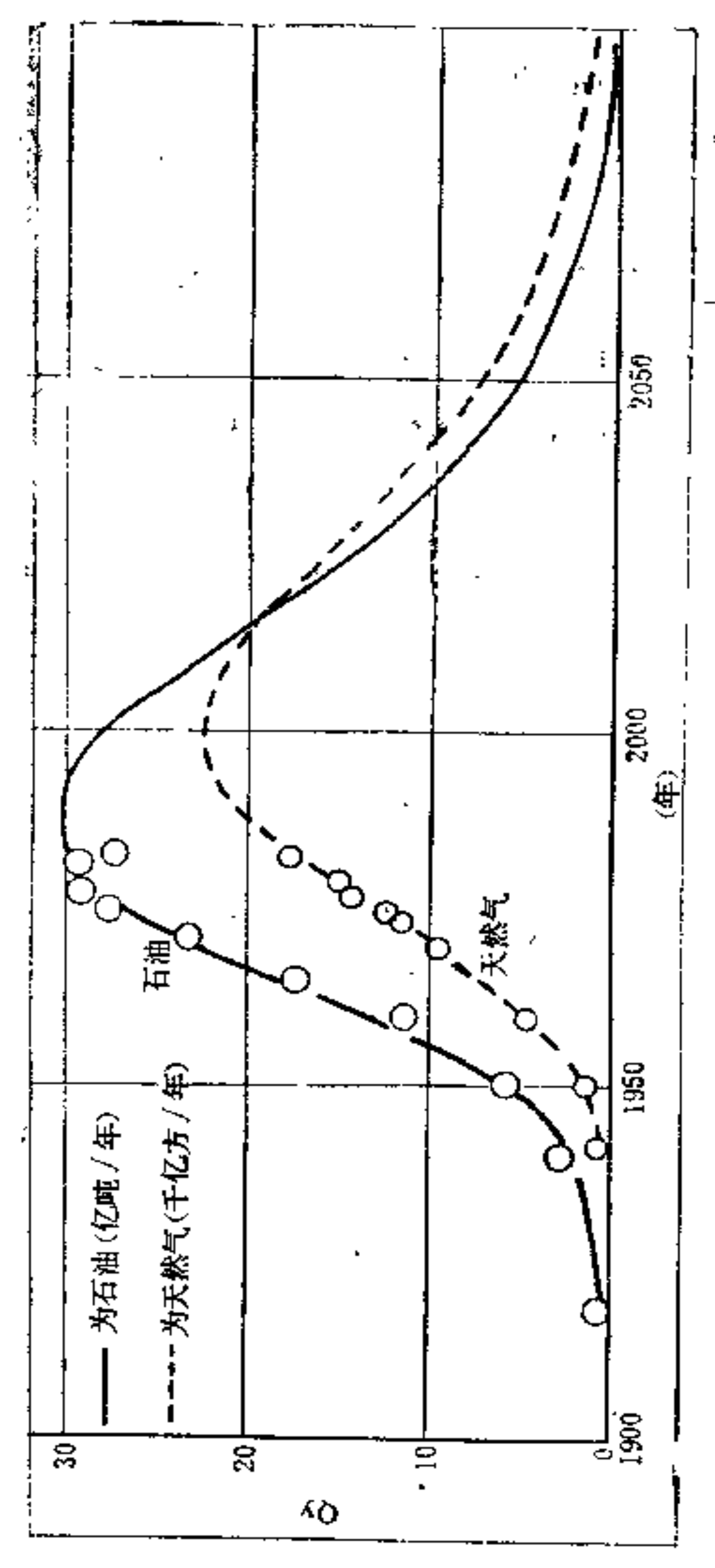


图 6—4 世界石油、天然气产量基值

从上式得出的盛衰曲线如图6—4所示。图6--4中实

线为 Q_y 值。大圆圈代表实际年产量。

从基值模型中，可以分出石油产量发展的四个概括性阶段：

1)加速上升阶段：1918～1962(年)

2)一段上升阶段：1962～1988(年)

3)一般下降阶段：1988～2014(年)

4)缓慢下降阶段：2014～2100(年)

截止1978年，已采出储量的模型估计值为

$$\sum_{y=1918}^{1978} Q_y \cong 530 \quad 亿吨$$

比实际采出量555·3亿吨①低4·5%。

到2100年的最终可采储量模型估计值为：

$$\sum_{y=1918}^{2100} Q_y \cong 2000 \quad 亿吨$$

比1978年的一种估计①555.3＋743.9＝1299.2亿吨高53%。比1980年巴黎召开的世界地质会等的估计②2570亿吨低22%。比十一届世界石油会议上的一种估计2460亿吨低19%③。

① 甘克文等，《世界含油气盆地图集》，石油工业出版社，1982.

② Wolf. Häfele et al., Energy in a finite wored, path to a sustainable future, Ballinger publishing Co. Cambridge, Massachusetts, publishers, Inc.1981

③ Charles D.Masters et al., Distribution and quantitative assessment of world petroleum reserve and resources, 11th world petroleum Congress, London, 1983

用上述模型估计1985年的年产量为：

$$Q_{1985} \cong 29.8\text{亿吨/年}$$

比卡拉尔①的估计值35·8亿吨/年低17％

又可估计1990年的产量为：

$$Q_{1990} \cong 30.0\text{亿吨／年}$$

也比卡拉尔的估计38.02亿吨／年低21％

以上基值也低于格兰斯瓦②的预测。可以认为这些差别是在世界政治因素影响的幅度以内的。

在本例中，从1962年（$t=4.4$）到1978年（$t=6$），产量翻了一番，即 $Q_{t2}/Q_{t1}\cong 2$，由此可以求得：

$$n \cong \frac{(6-4.4)+\ln 2}{\ln\left(\frac{6}{4.4}\right)} \cong 7.4$$

由于采收率可以大幅度提高，新油田还将不断发现，将来的实际年产量会比图6—4中曲线所代表的基值大，在下一世纪更是如此。

泊松旋回例二—— 世界天然气产量基值的宏观预测

和世界石油产量基值的宏观预测相似，在某些条件固定在80年代水平的假设下，世界天然气年产量基值

① R. E. Carlile, Seismic and drilling exploration, wordwide trends 1981—1985.
CGS-SEG 1981年 北京学术讨论会论文.1981.

② Manfred Grathwohl, Word energy Supply, resousces, technologies, Perspectives, Walter de Gruyter, Berlin, 1982 pp49—50, 1982.

(Q_y)的模型为：

$$Q_y = 1.3 + 0.028t^7e^{-t} \quad （千亿方／年）$$

$$t = (y - 1930)/10$$

式中y是公元年份。

预测基值Q_y（千亿方／年）和实际值$\hat{Q}_y$（千亿方／年）①比较如下（表6—2）：

表6—2

y	$\dot{Q}_y$	$\hat{Q}_y$	y	Q_y
1940	1.3	0.87	1985	18.7
1950	1.8	1.851	1990	20.7
1960	4.3	4.483	1995	21.9
1970	9.7	10.376	2000	22.3
1973	11.6	12.303	2010	21.0
1975	12.9	12.653	2020	17.8
1977	14.2	13.606	2030	14.0
1978	14.8	14.837	2060	7.4
1980	16.0	15.95	2100	1.8

世界天然气产量基值的盛衰可预测为4个阶段：

1)加速上升阶段　　1930～1973（年）

2)　　上升阶段　　1973～2000（年）

① Jonathan David Aronson et al., Profit and Pursuit of energy, market and regulation, Westview Press, Boulder, Colorado, p85, 1982.

3)　下降阶段　　2000～2026（年）

4)缓慢下降阶段　　2026～2100（年）

截止1978年已采出的天然气模型估计值为

$$\sum_{1930}^{1978} Q_y \cong 230 \quad （千亿方）$$

比1978年实际累计产出量251千亿方低8%

估计到2100年天然气产出总量为：

$$\sum_{1930}^{2100} Q_y = 1626 \quad （千亿方）$$

比1978年估计的251＋653＝904 千亿方高 80%，比1980年巴黎召开的世界地质会议等的估计2954～2980千亿方[①]低45%。

天然气产量模型中包括有发散部分（1.3），它可能包括当前（二十世纪八十年代）尚未足够开发的非正则气藏（如煤成气藏），地下水、海下水和冻土中的水合天然气。可能还有继续在生成的生物气（如我国的长江气）。

有人估计：在世界现有天然气储量的664千亿方中，有20%是生物气[②]。天然气水合物则是在高压低温下形

---

① Wolf Häfele et al., Energy in a finite World, Path to a Sustainable future, Ballinger Publishing Co. Cambridge, Massachusetles, Publishers, Inc.1981.

② Dadley D. Rice et al., Bull. A. A. P.G.65 (1) p5—25, 1981.

成的。深层冻土和海底沉积中存在这种条件。

由此可预测，在二十一世纪中，发散部分的天然气产量将超过收敛部分的产量。

天然气产量基值预测曲线，用虚线绘在石油年产量基值预测图6—4中，小圆圈是实际天然气产量。可以预计，上述基值也将为实际产量所超过。

§5 逻辑斯谛旋回

逻辑斯谛旋回模型可写成下列形式：

$$x=\frac{A}{1+ae^{bt}}$$

如b>0，这个模型可以形象一个体系生命末期 $\lim\limits_{t\to\infty} x\to 0$ 的过程。

如b<0，这个模型可以形象一个体系发展到最后极限 $\lim\limits_{t\to\infty} x\to A$ 的过程。这时，这个公式又称为比尔(pearl)公式。

逻辑斯谛函数也可以写成其它形式，如：

$$\because \qquad x=\frac{A}{1+ae^{-bt}}=\frac{\left(\frac{A}{a}\right)e^{bt}}{1+\frac{1}{a}e^{bt}}$$

当 A =1时，可以变换成：

$$\frac{x}{1-x} = \frac{1}{a} \cdot e^{bt}$$

在本世纪七十年代，John Fisher 等人（1970）用了一个类似的经验公式：

$$\frac{x}{1-x} = \frac{A}{a} e^{bt}$$

来形象新陈代谢或推陈出新的过程，X表示同类事物中后期事物代换前期事物的比率，t表示时间。所以这一公式也称为代换函数。实际应用时，当新产品已初步进入市场后，就能根据上式预测未来的代换过程。但是，从公式推论：完全代换是达不到的，显然，这种推论不适用于许多实际过程。

逻辑期谛旋回例一 —— 油层注水末期的油水比

某油田的一个试验区，油水比已降到10%以下，逐年平均油水比随着时间的下降关系可用下式拟合：

$$x = \frac{1}{1+11.3e^{0.186t}}$$

$$t = y - 1970$$

其中 y 表示公元年份。

拟合曲线x见图6—5。圆圈是平均油水比的实际值。

逻辑斯谛旋回例二——油层水油比和采收率

某油田的试验区中，油层有效孔隙率22%，有效渗透率240毫达西，到注水采油末期，估计采收率x和水

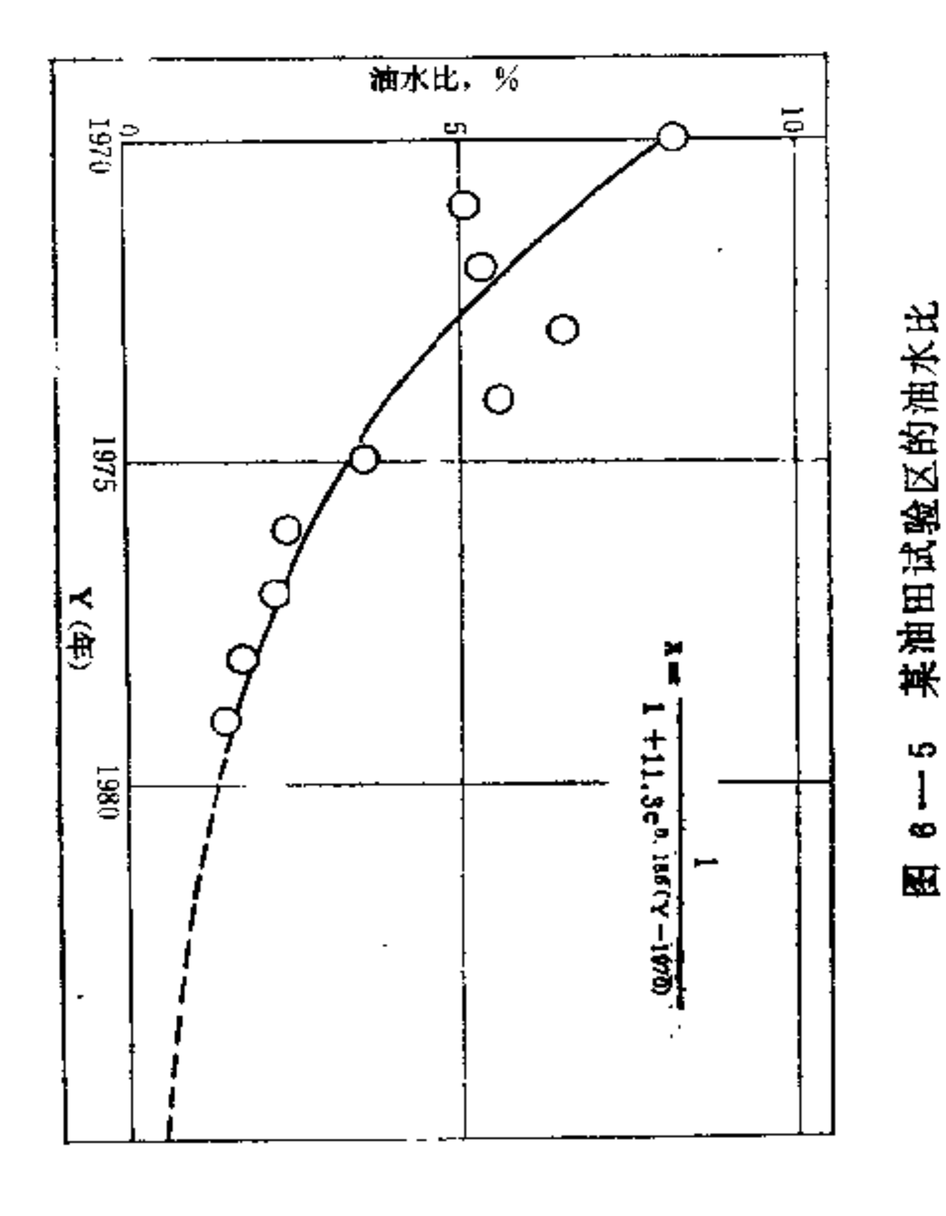


图 6—5 某油田试验区的油水比

油比 t 的关系可用下式拟合：

$$x = \frac{0.608}{1+0.4e^{-0.0116t}}$$

拟合曲线见图6—6。采收率极值约为60.8%。

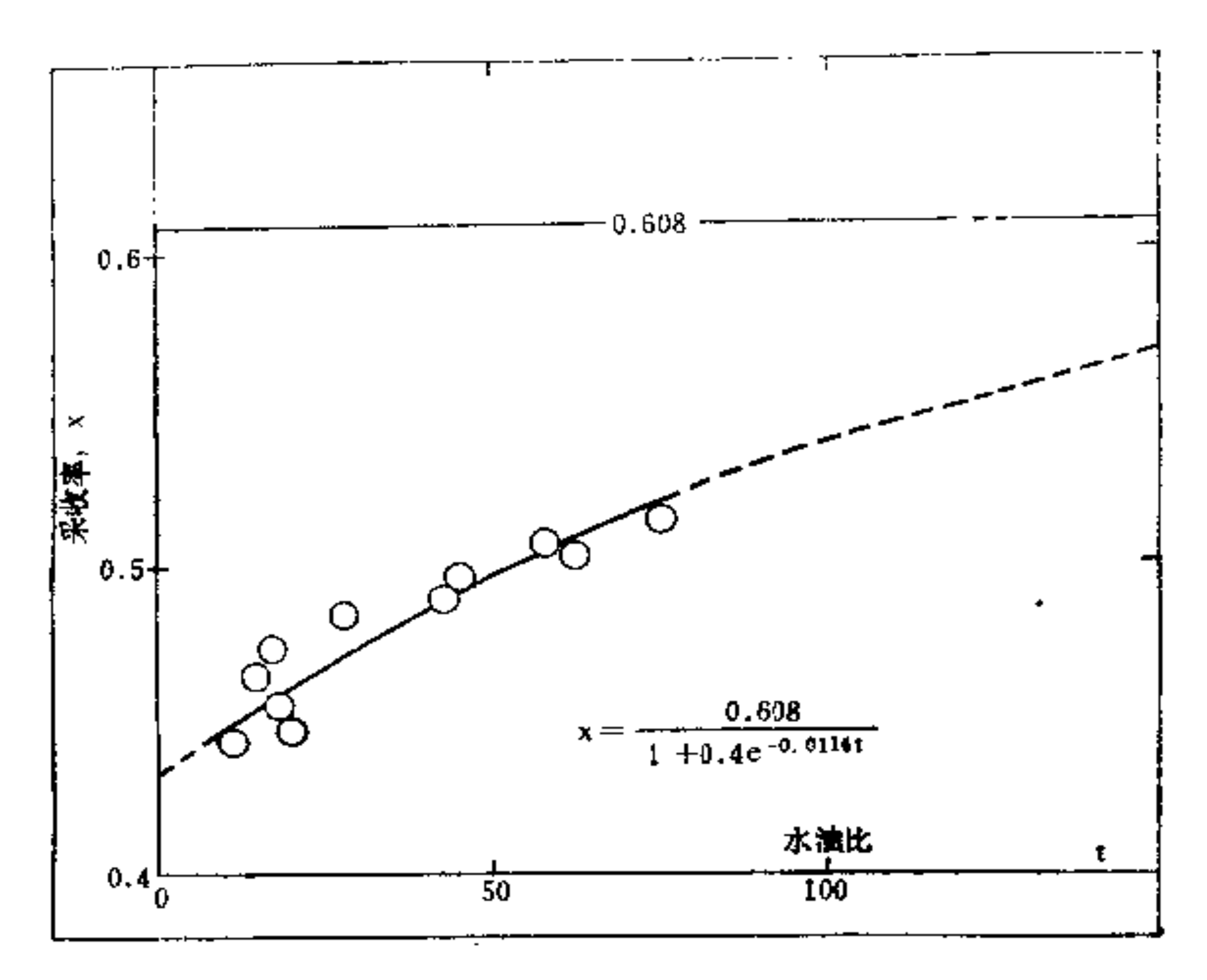


图6—6　某油田试验区采收率和水油比

第七章　回　　归

回归是统计学中的一个内容，也可以看作是根据最小二乘法原理同简单方程去拟合实际观察值的一种建立确定模型的方法。由于有两种不同观点，误差估计的方式也就不同。作为确定模型，可用均方根误差、平均绝对偏差（MAD）等表示。作为随机模型，在一定假设下，可用剩余平均方差，预测值估计方差等表示，也还可以估计预测值的置信区间。

回归预测方法是比较成熟的方法，已在技术经济预测中得到了广泛的应用，这里不再作详尽的讨论。

§1 自 回 归

对于序列〈x_i〉，$i=1, 2, \cdots, n$。τ阶自回归模型（有时记作AR（τ））可写成：

$$x_i=a_0+a_1x_{i-1}+a_2x_{i-2}+\cdots+a_\tau x_{i-\tau}$$

式中$a_0, a_1, \cdots, a_\tau$等是常系数。

上式的几种特款有：

1）如$a_0=a_1=a_2=\cdots=a_{\tau-1}=0$，$a_\tau=1$，则有：

$$x_i = x_{i-\tau}$$

τ 即为序列〈x_i〉的下标周期。

如果将〈x_i〉写成〈$x(t)$〉，则有：

$$x(t) = x(t-\tau)$$

如果 t 是时间，τ 就是时间周期。

2）如$\tau=1$，$a_1=1$，$a_2=a_3=\cdots=0$，则有：

$$x_i = a_0 + x_{i-1}$$

或 $$x_i - x_{i-1} = a_0$$

上式是〈x_i〉的一阶差分。因为a_0为常数，所以〈x_i〉称为一阶差分齐次——等差序列。当然，〈x_i〉也可以是高阶差分齐次序列。

3）如$\tau=1$，$a_2=a_3=\cdots=0$，则有：

$$x_i = a_0 + a_1 x_{i-1}$$

这就是回归趋向模型。

4）如假定〈x_i〉的平均值$M(x_i)$是一个常数，也就是说平均值守恒，即有：

$$M(x_i) = a_0 + M(x_i)(a_1 + a_2 + \cdots + a_\tau)$$

成 $$M(x_i) = \frac{a_0}{1-(a_1+a_2+\cdots+a_\tau)}$$

通常情况下，$M(x_i)$是一个大于零的常量，所以，自回归函数$AR(\tau)$存在平均值的一个必要条件是：

$$a_1 + a_2 + \cdots + a_\tau < 1$$

5）如果序列〈x_i〉是零和序列‘或已从〈$\bar{x}_i$〉化成

零和序列，$M(x_i)=0$ 即有

$$x_i=\tilde{x}_i-M(\tilde{x}_i)$$

$$a_0=0$$

这种序列相对于参量τ的自相关函数（在随机模型中称为协方差$Cov.(x_i, x_{\tau+i})$）为：

$$r=\frac{1}{n-\tau}\sum_{i=1}^{n-\tau}(x_i\ \ x_{\tau+i})$$

相应的τ级二阶矩自相关系数是：

$$\rho_\tau=\frac{\sum_{i=1}^{n-\tau}(x_i\cdot x_{\tau+i})}{\sqrt{\left(\sum_{i=1}^{n-\tau}x_i^2\right)\left(\sum_{i=1}^{n-\tau}x_{\tau+i}^2\right)}}$$

上两式已见于“乘法和乘法表”一节。

令$\tau=1, 2, \cdots(\tau-1), \tau$，得到：$\rho_1, \rho_2, \cdots\rho_{\tau-1}, \rho_\tau$。用尤尔—华尔格（Yule—Walker）方程可估计常系数：$a_1, a_2, \cdots, a_\tau$：

$$\begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{\tau-2} & \rho_{\tau-1} \\ \rho_1 & 1 & \cdots & \rho_{\tau-3} & \rho_{\tau-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho_{\tau-2} & \rho_{\tau-3} & \cdots & 1 & \rho_1 \\ \rho_{\tau-1} & \rho_{\tau-2} & \cdots & \rho_1 & 1 \end{pmatrix}\begin{pmatrix} a_1 \\ a_2 \\ \cdots \\ a_{\tau-1} \\ a_\tau \end{pmatrix}=\begin{pmatrix} \rho_1 \\ \rho_2 \\ \cdots \\ \rho_{\tau-1} \\ \rho_\tau \end{pmatrix}$$

§2 线性回归

设有一一对应的零和序列：

$$\langle x_i \rangle = \langle x_1, x_2, \cdots, x_i, \cdots x_n \rangle$$

$$\langle t_i \rangle = \langle t_1, t_2, \cdots, t_i, \cdots t_n \rangle$$

假设它们的线性关系是：

$$x_i = a + bt_i$$

用最小二乘法可估计参量a、b 和相关系数 r：

$$b = \sum_{i=1}^{n} x_i \quad t_i / \sum_{i=1}^{n} t_i^2$$

$$a = \frac{1}{n}\sum_{i=1}^{n} x_i - \frac{b}{n}\sum_{i=1}^{n} t_i$$

$$r = \frac{\sum_{i=1}^{n} x_i \quad t_i}{\sqrt{\left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} t_i^2\right)}}$$

式中r是互相关系数。如将第五章互乘表一节的最后一个公式中的 d, g, m 改为 x, t, n 即得上式。

有许多函数可化成线性回归①。其中特别重要的一种形式是：

$$x = a + b(y - y_0)^2$$

如令：

$$x = \ln Q$$

$$a = \ln \frac{A}{\sqrt{2\pi}}$$

$$b = -\frac{1}{2\sigma^2}$$

① 《数学手册》，人民教育出版社，842页，1979。

式中A、σ是常参量，将Q化为x的函数即得：

$$Q=A\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{y-y_0}{\sigma}\right)^2}$$

$$=A\quad f\left(\frac{y-y_0}{\sigma}\right)$$

式中$f\left(\frac{y-y_0}{\sigma}\right)$是正态密度函数。

这就是上章所说的正态旋回模型。

逻辑斯谛旋回式是另一种可以化为线性回归的例子：

$$x=\frac{A}{1+ae^{bt}}$$

上式可写成：

$$\ln\left(\frac{A}{x}-1\right)=\ln a+bt$$

用适当的方法，利用现成的线性回归子程序，可以从〈(x_i, t_i)〉，i=1，2，…，快速估计a和b。

还有一种常用的化为线性回归的形式，就是将变量成应变量化为对数值。这种形式，我们暂用含有广泛含义的名词“同态线性回归”去形容它。

化成线性回归例一——北京市蛋和肉的生产

为了增强北京市民的体质，有关科技人员为增加动

物蛋白质的生产，作了各种因素的线性规划。蛋和肉是规划中两项动物蛋白质的重要来源。为验证规划的可行性，根据1970～1982年内的历史统计资料作试性回归预测。鸡蛋年产量E和生猪存栏量P的一种预测模型①如下：

$$E=933+107t_1 \qquad t_1=(y-1974)^2$$

$$P=239-1.62t_2 \qquad t_2=(y-1978)^2$$

其中 y 是公元年份。

E和P值随y而变的曲线见图7—1。由图7—1可以看

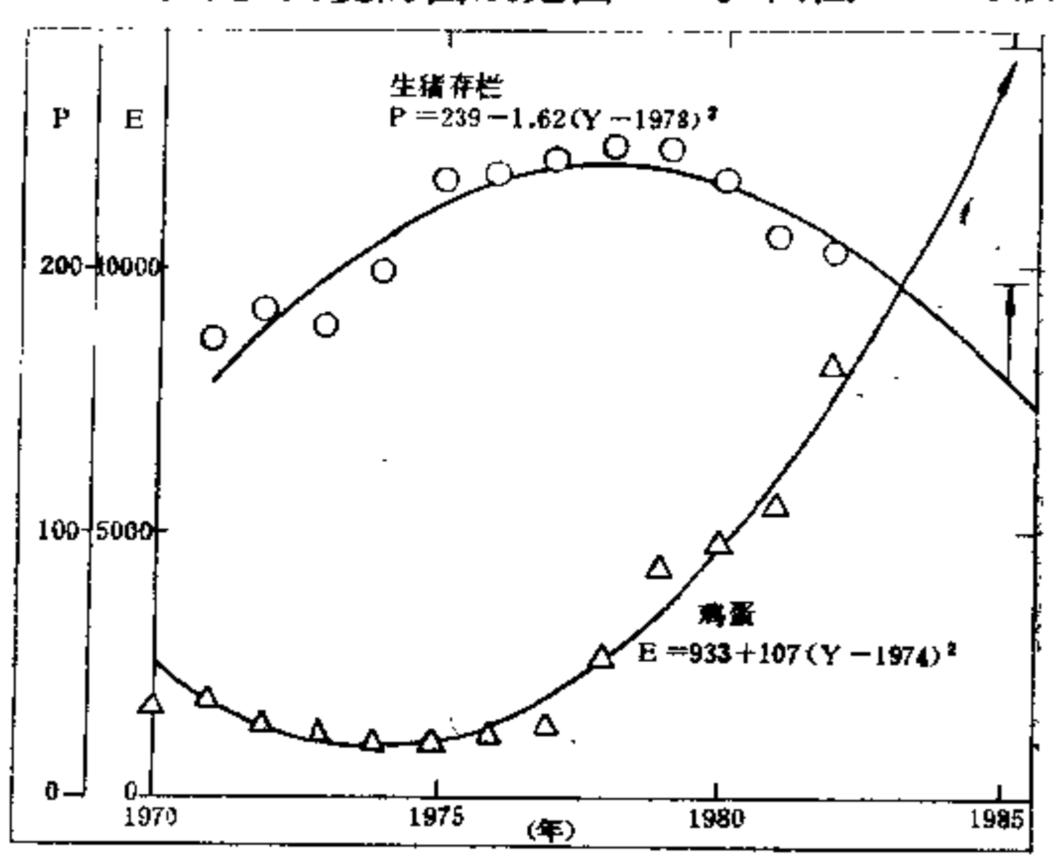


图7—1 北京市生猪存栏和鸡蛋产量

① 孙振玉、翁心林等，对目前京郊动物蛋白生产结构的调整意见，《科技参考》，北京市科学技术情报研究所，41（5），52页，1983.

出，到1985年P值下降很大，应当采取措施使产量维持在“理想”水平，如图7—1中箭头所示。

以上预测不但是很短期的，而且不会很准确。市民未必以每一货币单位能购买最多的动物蛋白质的原则上市。其中一定有许多人要考虑胆固醇等其它生化条件，或单纯为了好吃上市。这些情况当然也会影响发展趋势。

化为线性回归例二——估计世界石油的储产比

世界石油年产量Q_y与已探明的剩余可采储量Q_r之间，有一个大致的比例关系，Q_r是过去探明的可采储量加上当年探明的可采储量再减去当年采出的产量。可见Q_r包含当年从探到采的动态反情因子。

根据1950～1980年这30年内从几方面估计的以陆上油田为主的Q_r值，利用线性回归公式

$$\ln\left(\frac{A}{Q_r}-1\right)=\ln a+bt$$

可以求得逻辑斯谛旋回式：

$$Q_r=\frac{950}{1+42.1e^{-0.166t}}$$

$$t=y-1940$$

式中y是公元年份。

y对Q_r的关系见图7—2。图中曲线是上式的计算值，圆圈是实际值。Q_r的临时上限估计为950亿吨。

又从1950～1980年这30年内石油年产量Q_y和剩余探

明可采储量 Q_r 的资料，可以粗略地作一次线性回归分析。Q_r/Q_y 比值对公元年份 y 的线性回归式是：

$$Q_r/Q_y \cong 38 - 0.26(y - 1950)$$

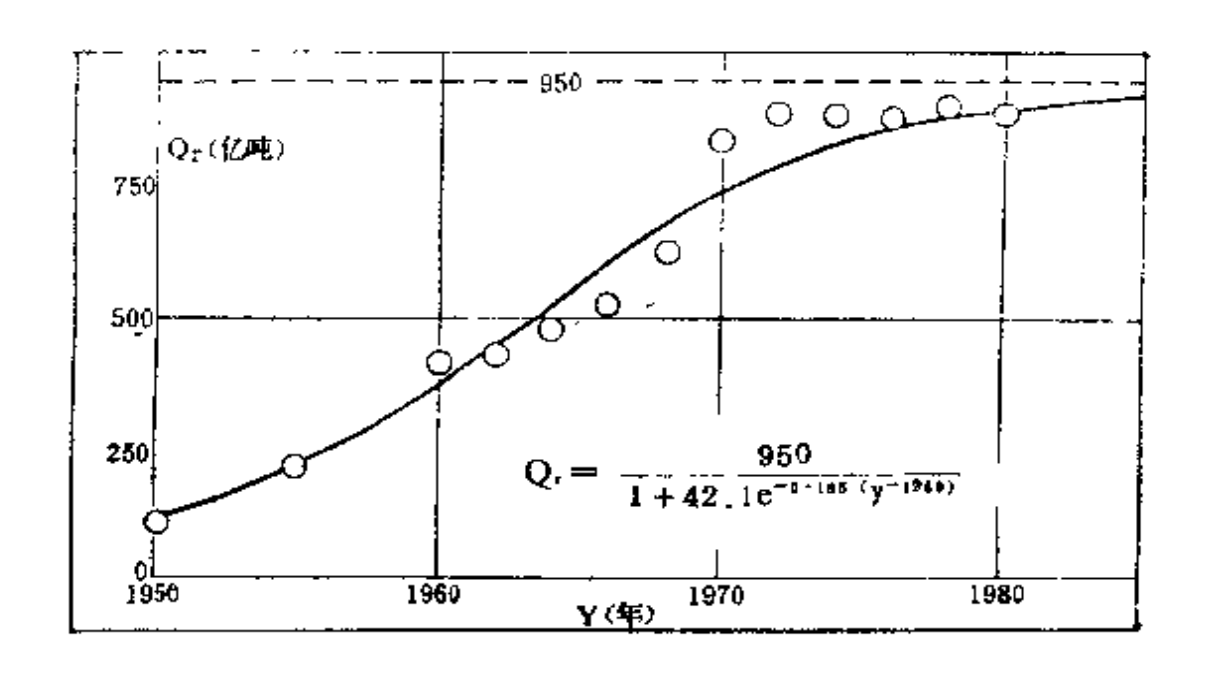


图 7—2　世界石油已探明的剩余可采储量

计算结果如图7—3。圆圈代表实际值。

由以上回归分析可以预测 1990 年和 2000 年的 Q_r/Q_y 比值。

$$\left(\frac{Q_r}{Q_y}\right)_{1990} \cong 27.6$$

$$\left(\frac{Q_r}{Q_y}\right)_{2000} \cong 25$$

Q_r/Q_y 比值下降的原因可能是低储采比油田比重增加所引起的。

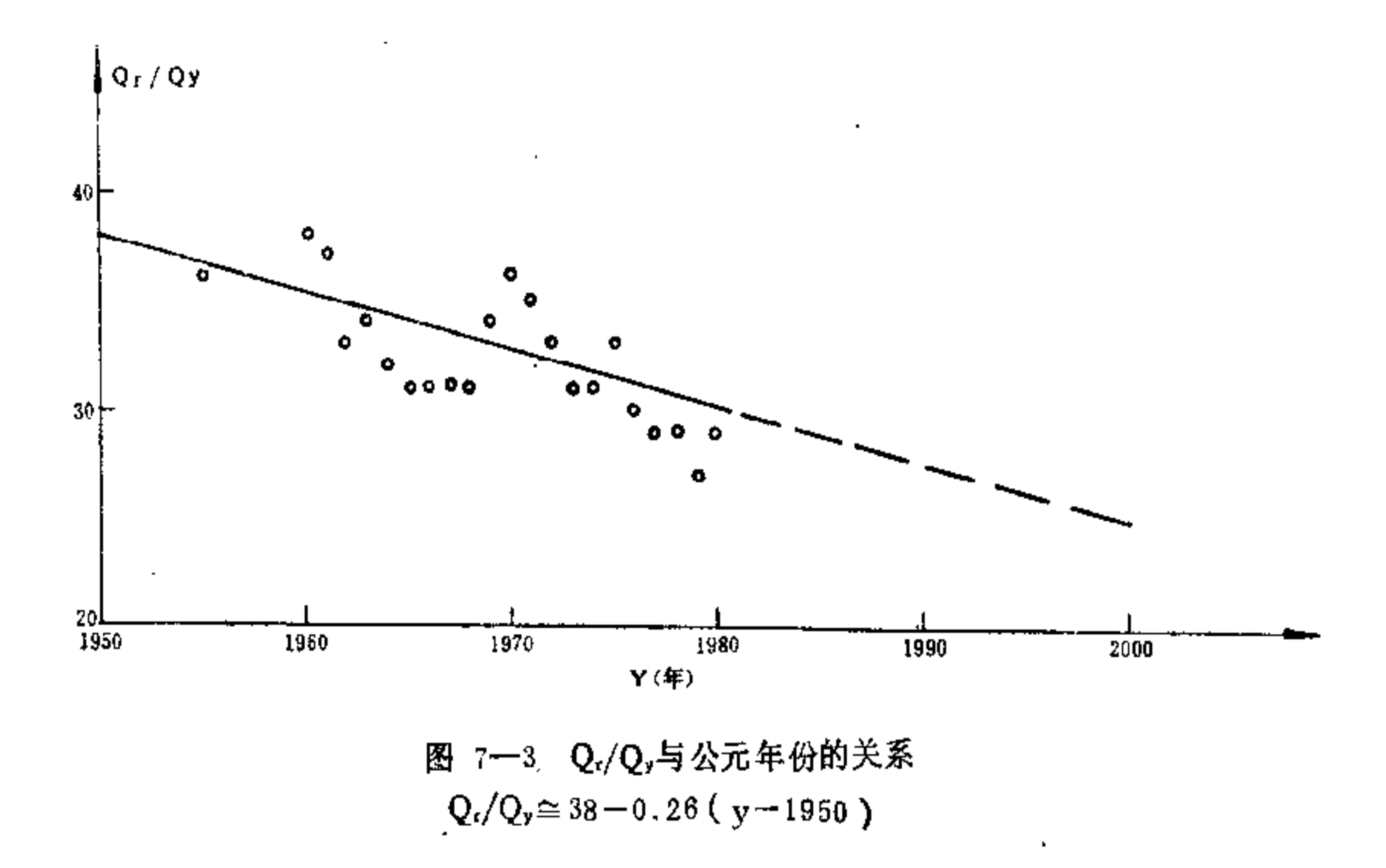


图 7—3 Q_r/Q_y与公元年份的关系

$Q_r/Q_y \cong 38-0.26(y-1950)$

§3 同态线性回归

线性回归的广泛应用表明：许多客观体系常常可以用一段直线来形象某一个局部关系。扩张到它的同态（或称同坯）关系也可以用来形象某些其它体系中存在的关系。其中最重要的是对数形式的线性回归：

1） $\ln x = a + bt$

即 $x = (e^a) \cdot e^{bt}$

2） $\ln x = \ln x_0 + t \cdot \ln(1+\alpha)$

即 $x = x_0 (1+\alpha)^t$

第二式在拟合信号一章中已提到过，称为复利函数。α 常表示国家或企业计划的年增加率、银行存款利率、贴观观金回收利润率（DCFR）、人口出生率和人口自然增长率等。如果令

$$b = \ln(1+\alpha)$$

则 $$\alpha = e^b - 1$$

参量 α 又可称为综合递增率。因为综合递增率可用最小二乘法最优化来估计，一般比平均递增率更有代表性。

同态线性回归例一——我国人口自然增长预例

近年来，我国人口 x 的自然增长可以用对数线性回归式来形象：

$$\ln x = 0.019204y - 26.5238$$

式中x是公元y年年底的人数，以万人为单位，y是公元年份。表7—1是计算值x和实际调查值$\hat{x}$的比较。

表7—1　计算值x和实际调查值$\hat{x}$①

y	x（万人）	$\hat{x}$（万人）	$\hat{x}-x$（万人）	y	x（万人）	$\hat{x}$（万人）	$\hat{x}-x$（万人）
1949	54437	54167	−270	1981	100642	100072	−570
1952	57665	57482	−180	1982	102593	101541	−1052
1957	63477	64653	+1176	1985	108677		
1965	74018	72538	−1480	1990	119630		
1978	95008	96259	+1251	1995	131683		
1979	96850	97542	+692	2000	144958		
1980	98728	98705	−23				

① 国家统计局，《中国统计摘要》，中国统计出版社，13页，1983。

y和$\ln\hat{x}$的二阶矩相关系约99.89%，由表7—1可见，如不设法控制人口，到2000年人口将远超过12亿人。

上式中的参量$b=0.019204$可换算为人口自然增加平均年率$\bar{\alpha}=e^b-1\cong1.94\%$。这比1982年的实际值1.45%高得多。如果用最低的1979年的实际值1.17%外推，到2000年人口约为13.2亿，仍超过12亿人。

对于采取措施控制生育后，人口下降的条件，平均生育年龄的影响等问题都是当前在研究的问题，可参看

近期有关出版物。有一种估计认为，如1986年做到只生一胎，到2004年人口才能下降。

上述对数线性回归模型是以拟合程度最高为原则的。本质上是一种纯粹的唯象模型。如果考虑到事物发展阶段（“S”阶段），许多体系可用正态密度函数拟合（如正态旋回例一和例二），那么可假定1982年采取强烈措施，使人口由猛升阶段人为地纳入上升阶段，这样就可得出第二模型：

$$\ln x = 2.83 - \frac{(2050-y)^2}{9000} \qquad \text{（亿人）}$$

上式的计算值x和实际值$\hat{x}$比较如下（表7—2）：

表7—2

y	x（亿人）	$\hat{x}$（亿人）	（$\hat{x}-x$）（亿人）	y	x（亿人）
1949	5.46	5.42	−0.04	1985	10.60
1952	5.83	5.75	−0.08	1990	11.36
1957	6.48	6.47	−0.01	1995	12.11
1965	7.59	7.25	−0.34	2000	12.84
1978	9.53	9.63	+0.10	2010	14.15
1979	9.68	9.75	+0.07	2020	15.33
1980	9.83	9.87	+0.04	2030	16.21
1981	9.98	10.00	+0.02	2040	16.76
1982	10.14	10.15	+0.01	2050	16.95

① 卢泽愚：种群增长的矩阵计算模型，生态学报，1（3），1981。

这个模型引入了信息的相似原则，所以可以外推得远一些。在1949～1982年期间，拟合程度不如第一模型，特别是1965年偏差（$\hat{x}-x$）较大。但总的符合程度仍能说明第二模型并未显著背离历史发展实际。因此是可能成立的。根据第二模型，中国人口到二十一世纪中期达到饱和，人数约16～17亿人。如果将人口饱和点提前在2000年初到达，就要采取严格的措施。

同态线性回归例二——中国社会总产值的基值预测

中国近年来的社会总产值x，可用同态回归形象：

$$\ln x = -3.8043 + 0.074051(y-1900) \quad (\text{千亿元})$$

式中x是公元y年的社会总产值，以千亿元人民币为单位。

计算值x和实际值$\hat{x}$①比较如下（见表7—3）：

表7—3

y	x（千亿元）	$\hat{x}$（千亿元）	$\hat{x}-x$（千亿元）	y	x（千亿元）	$\hat{x}$（千亿元）	$\hat{x}-x$（千亿元）
1952	1.047	1.015	−0.032	1981	8.970	9.048	+0.078
1957	1.517	1.606	+0.089	1982	9.659	9.894	+0.235
1965	2.743	2.695	−0.048	1985	12.062		
1978	7.183	6.846	−0.337	1990	17.467		
1979	7.735	7.642	−0.093	1995	25.293		
1980	8.329	8.496	+0.167	2000	36.628		

y和$\ln\hat{x}$的二阶矩相关系数约99.92%。由表7—3可

① 国家统计局：《中国统计摘要》1983，中国统计出版社，5页，1983。

见，预测2000年的社会总产值为1680年的4倍，也就是说，只要解决好关键问题，那末，翻两番的目标是可以达到的。表7—3中（$\hat{x}-x$）的负值表示社会总产值退步，正值表示社会总产值进步。

从回归式可估计综合年增长率为：

$$\alpha = e^{0.07405} - 1 = 7.69\%$$

同态线性回归例三——中国粮食生产基值预测

中国近年来粮食生产基值x可用同态线性回归形象：

$$\ln x = 0.0288(y-1939)（亿吨）$$

计算值x和实际值$\hat{x}$ 比较如下：

表7—4

y	x（亿吨）	$\hat{x}$（亿吨）	$\hat{x}-x$（亿吨）	y	x（亿吨）	$\hat{x}$（亿吨）	$\hat{x}-x$（亿吨）
1949	1.33	1.13	−0.20	1978	3.07	3.05	+0.02
1952	1.45	1.64	+0.19	1979	3.16	3.32	−0.16
1957	1.68	1.95	+0.27	1980	3.26	3.21	−0.05
1965	2.11	1.95	−0.16	1981	3.35	3.25	−0.10
1975	2.82	2.84	+0.02	1982	3.45	3.53	+0.08

从回归式可估计综合年增产率为：

$$\alpha = e^{0.0288} - 1 = 2.9\%$$

这个数值远比人口自然增长率高。因为农业生产是有极值的，上列线性回归式不能远程外推。

①国家统计局，《中国统计摘要》，中国统计出版社，25页，1983，

§4 多元回归

设有 n 组 l 元数据：

$$y_i = y_i\ (t_1,\ t_2,\ \cdots,\ t_j,\ \cdots,\ t_l)$$

即有：

$$y_1 = y_1\ (t_{11},\ t_{12},\ \cdots,\ t_{1j},\ \cdots,\ t_{1l})$$

$$y_2 = y_2\ (t_{21},\ t_{22},\ \cdots,\ t_{2j},\ \cdots,\ t_{2l})$$

………

$$y_i = y_i\ (t_{i1},\ t_{i2},\ \cdots,\ t_{ij},\ \cdots,\ t_{il})$$

………

$$y_n = y_n\ (t_{n1},\ t_{n2},\ \cdots,\ t_{nj},\ \cdots,\ t_{nl})$$

假设用下式拟合：

$$y_i = a_0 + a_1 t_{i1} + a_2 t_{i2} + \cdots + a_j t_{ij} + \cdots + a_l t_{il}$$

用最小二乘法，可求得参量 a_0，a_1，…，a_l的估计式为：

$$\begin{pmatrix} n & \Sigma t_{i1} & \Sigma t_{i2} & \cdots & \Sigma t_{ij} & \cdots & \Sigma t_{il} \\ \Sigma t_{i1} & \Sigma t_{i1}^2 & \Sigma t_{i2}\cdot t_{i1} & \cdots & \Sigma t_{ij}\cdot t_{i1} & \cdots & \Sigma t_{il}\cdot t_{i1} \\ \Sigma t_{i2} & \Sigma t_{i1}\cdot t_{i2} & \Sigma t_{i2}^2 & \cdots & \Sigma t_{ij}\cdot t_{i2} & \cdots & \Sigma t_{il}\cdot t_{i2} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \Sigma t_{ij} & \Sigma t_{i1}\cdot t_{ij} & \Sigma t_{i2}\cdot t_{ij} & \cdots & \Sigma t_{ij}^2 & \cdots & \Sigma t_{il}\cdot t_{ij} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \Sigma t_{il} & \Sigma t_{i1}\cdot t_{il} & \Sigma t_{i2}\cdot t_{il} & \cdots & \Sigma t_{ij}\cdot t_{il} & \cdots & \Sigma t_{il}^2 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \cdots \\ a_j \\ \cdots \\ a_l \end{pmatrix} = \begin{pmatrix} \Sigma y_i \\ \Sigma t_{i1} y_i \\ \Sigma t_{i2} y_i \\ \cdots \\ \Sigma t_{ij} y_i \\ \cdots \\ \Sigma t_{il} y_i \end{pmatrix}$$

式中符号“Σ”是 $\sum_{i=1}^{n}$的简写。

第八章 随机体系

随机性的概念可以看作是确定性概念的扩张。根据随机性的概念，一个随机事件A在一定条件下可能发生，也可能不发生，事件A发生的可能性用概率P(A)，$0 \leq P(A) \leq 1$ 来表示。确定模型就是随机模型在 $P=1$ 的假设下的特款。随机模型的演算比确定模型烦琐得多，例如两个独立随机变量的和，其概率密度函数就要用褶积来表示。

在随机根型中，均匀分布系和正态分布系有特别重要的意义。均匀分布系只有一个主参量，那就是平均值，或者概率。除一般的均匀分布外，二项分布、泊松分布都属于均匀系。正态分布系有二个主参量，它们是平均值和方差。（第六章中已提到过泊松概率函数和正态概率密度函数都是从确定性模型中引出的确定性函数，并没有联系到有关的随机性）。无论是均匀分布系，或是正态分布系，平均值都是一个重要的参量。技术经济预测中的平均法及其衍生的预测法都是以平均值守恒或相对守恒为基础的。

§1 分　布

对于确认为一元随机性的数据体系，如半序集〈x_i〉$i=1, 2, \cdots, n$. 可假设下列"分布函数"作为模型：

$$0 \leqslant F(x) \leqslant 1$$

式中F(x)是随x单调增加的连续、可导函数。

从实际数据中，可求得与分布函数$0 \leqslant F(x) \leqslant 1$对应的经验分布$\hat{F}(x)$，它有如下几种表示法：

直接式　$\hat{F}(x)=\frac{i}{n}$　　$\frac{1}{n} \leqslant \hat{F}(x) \leqslant 1$

无偏式一　$\hat{F}(x)=\frac{2i-1}{2n}$

$$\frac{1}{2n} \leqslant \hat{F}(x) \leqslant \frac{2n-1}{2n}$$

无偏式二　$\hat{F}(x)=\frac{i-1}{n-1}$　　$0 \leqslant \hat{F}(x) \leqslant 1$

作为模型的分布函数F(x)对x求导数，得到概率密度函数(p.d.f)：

$$f(x)=\frac{d}{dx}F(x)$$

对应于f(x)的经验概率密度可用下式表示：

$$\hat{f}(x)=\frac{\Delta i}{n}$$

§2 分布的数字特征

分布的数字特征是反映分布性质的统计量。

设经验分布为半序集（x_i） $i=1, 2, \cdots, n$。它的模型的分布函数为F(x)，概率密度函数为f(x)。其数字特征为：

1）数学期望M_x：符号M表示平均值，或以概率密度函数为权的积分值。

对于数据经验分布有：

$$\hat{M}_x=\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$$

对于模型分布函数有：

$$M_x=\int x\cdot f(x)dx$$

数学期望 M_x 也称为一阶原点矩。k阶数据经验分布原点矩是：

$$\hat{M}(x^k)=\frac{1}{n}\sum_{i=1}^{n}x_i^k$$

对模型分布函数是：

$$M(x^k)=\int x^k\cdot f(x)dx$$

2）方差 D_x：

对数据经验分布有：

$$\hat{D}_x = \hat{M}(x-\bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

对模型分布函数有：

$$D_x = \int (x-\bar{x})^2 f(x)\,dx$$

方差 D_x 又称二阶中心矩。以 $\bar{x}$ 为中心数据经验分布的k阶中心矩是：

$$\hat{M}(x-\bar{x})^k = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^k$$

对模型分布函数是：

$$M(x-M_x)^k = \int (x-M_x)^k f(x)\,dx$$

3）中位值 a_x

对数据经验分布有：

$$\hat{a}_x = \begin{cases} x_{\frac{n+1}{2}} & n\text{为奇数} \\ \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n+2}{2}}\right) & n\text{为偶数} \end{cases}$$

对模型分布函数有：

$$a_x = F^{-1}\left(\frac{1}{2}\right)$$

式中 F^{-1} 表示 $F(x)$ 的逆函数。

§3 正态分布

很多达到平衡状况的体系，其元素分布接近于正态

分布。正态分布体系可作为随机体系的一个典型。

根据随机性的假定，如概率相交（高斯1974）、二项分布的扩张（拉普拉斯1812）或随机量的等概率组合（汉根1837）等，可以推导出正态分布函数或它的特款。

正态分布函数是：

$$F(x)=\frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{x-a}{\sigma}\right)^{2}}dx$$

$$=N(a,\ \sigma)$$

概率密度函数是：

$$f(x)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{1}{2}\left(\frac{x-a}{\sigma}\right)^{2}}$$

其数字特征为：

$$M_x=a_x=a$$

$$Dx=\sigma^2$$

k 阶原点矩为：

$$M(x^k)=\begin{cases}a & k=1\\ \sigma^2+a^2 & k=2\\ a(a^2+3\sigma^2) & k=3\end{cases}$$

如果以正态分布作为下列实际经验分布的模型：

$$\langle \hat{x}_i \rangle=\langle x_1,\ x_2,\ \cdots,\ x_i,\ \cdots,\ x_n \rangle$$

则有参量估计式：

$$a=\hat{M}_x=a_x=\frac{1}{n}\sum_{i=1}^{n}x_i$$

$$\sigma=\sqrt{\hat{D}_x}=\left[\frac{1}{n}\sum_{i=1}^{n}(x_i-a)^2\right]^{\frac{1}{2}}$$

如以“$\frac{1}{n-1}$”代替上式中的“$\frac{1}{n}$”就得到相应的无偏值。

统计手册中一般都给出了标准正态分布函数$\Phi(u)$和标准正态概率密度函数$\varphi(u)$的值:

$$\Phi(u)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{u}e^{-\frac{u^2}{2}}du$$

$$\varphi(u)=\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}$$

如果有了估计参量a, σ，就可以通过u对x的标准化变换:

$$u=\frac{x-a}{\sigma}$$

从$\Phi(u)$和$\varphi(u)$的表中查到$F(x)$，$f(x)$随x而变的值。

正态分布$N(a,\sigma)$中l个元素的和的分布也是正态分布，分布函数是$N(la, \sqrt{l}\,\sigma)$。

平面极座标上的正态分布称为瑞利分布。瑞利分布的概率密度函数是:

$$f(\rho)=\frac{\rho}{\sigma^2}e^{-\rho^2/2\sigma^2}$$

$$\rho\geqslant0$$

数字特征有：

$$M\rho=\sqrt{\frac{\pi}{2}}\sigma$$

$$D_x=\frac{4-\pi}{2}\sigma^2$$

瑞利分布可作为二维白噪音振幅ρ的模型。

正态分布已得到广泛的承认和应用，正态分布的导出分布t一分布和F一分布等也得到了广泛的应用。当然，正态分布也还有新的问题需要研究，如受限制的正态分布，形式和标准正态密度相同的确定模型等问题。

在第七章线性回归一节中，对于数振序列：〈(x_i, t_i)〉，i=1,2,…,n，已给出了线性回归参量a，b和相关系数γ的估计式。通常都是假设实际数据与回归结果之间的误差服从正态分布。因此还可以估计相关系数γ的置信限γ_a，（见通用统计表[1]）。

正态分布例一——世界粮食单位产量

近10～20年来，世界上几个粮食生产国的主要粮食单位面积产量（斤/亩）和有关统计量[2][3]如下（见表

① 中国科学院数学研究所概率统计室，《常用数理统计表》，科学出版社，18页，114—116页，1974。

② 世界经济年鉴，1981。

③ 国家统计局，《中国统计摘要》，24—25页，1983。

8—1）：

表3 1

类别	y	n	$\hat{M}_x$	$\hat{a}_x$	x_v	Sup.x	N(a, σ)	D_n
小麦	1957	7			305		N(190, 90)	0.10
	1970	33	489	539	743	817	N(500, 190)	0.08
	1979	23	451	419	791		N(453, 233)	0.07
水稻	1960	10	194	180	261	283	N(194, 53)	0.10
	1970	10	251	232	313	353	N(253, 47)	0.15
	1979	24	349	547	847		N(560, 267)	0.10
玉米	1970	21	325	463	964		N(480, 243)	0.08
	1979	21	436	773	1036		N(587, 280)	

表中y是公元年份，n是数据个数（容量），$\hat{M}_x$是经验分布的平均，$\hat{a}_x$值是数据中位值，x_v是一国的实际最高值，Sup.x是大面积实际高产记最或称记录水平。N(a,σ)是根据数据所建立的正态分布函数的模型。D_n是正态分布函数和经验分布之间的最大离差。

小麦：$\hat{M}_x \sim \hat{a}_x \sim a$，符合正态分布。资料基本上反映出先导预测的原则，即：过去的记录水平（1970年817斤/亩）可能成为现在的先进水平（1979年791斤/亩）。如果现在的先进水平可能成为将来的平均水平，那么，我国小表单产（1982年约为327斤/亩）可能大幅度增加。

水稻：1970年前$\hat{M}_x \sim \hat{a}_x \sim a$，符合正态分布，1979年$\hat{a}_x \sim a$，但明显大于$\hat{M}_x$。可能有重大技术改革在推广之

中。我国水稻单产（1982年约650斤/亩）已超过世界平均值，接近先进水平（847斤/亩）。

玉米：$\hat{a}_x \sim a$，明显大于$\hat{M}_x$。也可能有重大技术改革在推广之中。如果今日的先进水平（$x_v = 1036$斤/亩）预示着将来可能达到的平均水平，那么，我国单产（1982年约434斤/亩）可能大幅度增加。

§4 平均法

平均值预测法，或简称平均法，是以“期望平均值守恒”为根据的。凡属于均匀分布和正态分布的体系都具备这一性质。

在均匀分布体系类中有：均匀分布体系，二项分布体系，泊松分布体系等。属于这类体系的自然模型有：产品的正品率，射击的命中率，运动员的得分数，机床运转率，放射源衰变数，细胞染色体交换率，原料或产品的缺陷率，电话交换次数，传动带上零件到达率，疫病发生率等。这些模型都可以用这类带有固定概率的分布函数预测。

在正态分布体系中，不但“期望平均值守恒”，并且“期望方差守恒”。所以正布分布模型可预测：量度的随机误差，射击偏差值，一些生物的长度、重量、出生率和死亡率，工农业原料或成品的重量、大小、物性

和成份，药物的生理反应，产品或设备的寿命等。

市场预测中，严格具备“期望平均值守恒”的情况是不多的。但现行的直观方法，如：顾客需要直接调查，经理、售货员、专家意见调查，市场因子的推演[1]等都期望在一定程度内，通过资料综合来消除随机干扰因素。

在定量预测中，简单的平均法（SA）是直接根据“期望平均值守恒”的一种预测方法。移动平均法则期望平均演在短期内守恒。加权移动平均法（WMA）和指数平滑法期望平均值在短期内格近或局部守恒。

§5 移动平均法

移动平均法是平均法的扩张。对于不同类型的体系，用移动平均法预测一个时间序列（x_i）=（$x_1, x_2, \cdots, x_i, \cdots, x_n$）的技术有下列几种：

1)算术移动平均

平均值：$M_i = \frac{1}{2k+1}\sum_{l=i-k}^{i+k} x_l$

$l=1, 2, \cdots, \quad k=1, 2, \cdots$

趋向值：$\Delta M_i = M_i - M_{i-1}$

预测值：$x^*_{i+l} = M_{i-k} + \Delta M_i(k+1)$

2)加权移动平均

① 郭军元：《市场学》，机械工业出版社，131—149页，1982。

平均值：$M_i=\sum_{l=i-k}^{i+k} a_l x_l \Big/ \sum_{l=i-k}^{i+k} a_l$

a_l是加权系数，如$k=1$时可以令$a_{i+1}=3$，$a_i=2$，$a_{i-1}=1$使近期数据有较大的权。

趋向值：$\Delta M_i=M_i-M_{i-1}$

预测值：$x^*_{i+1}=M_{i-k}+\Delta M_i(k+1)$

3）指数移动平均

平均值：$M_i=ax_i+(1-a)M_{i-1}$，　　$0\leqslant a<1$

a 接近于零表示平均值M_{i-1}有较大的权，a 接近于1表示最近值x_i有较大的权。比值$(1-a)/a$称为滞后系数。

趋向值：$\Delta M_i=M_i-M_{i-1}$

预测值：$x^*_{i+1}=ax_i+(1-a)M_{i-k}+(k+1)\cdot\Delta M_i$

§6 指数平滑法

对于数据序列$\langle x_r\rangle=\langle x_1,x_2,\cdots,x_i,\cdots,x_n\rangle$ 指数平滑法的递归式是：

$$x_1^*=x_1$$

$$x^*_{i+1}=ax_i+(1-a)x^*_i$$

$$0\leqslant a\leqslant 1$$

式中x^*_{i+1}，x^*_i都是预测值。“指数”a是权系数，反映了对新数据x_i的重视程度。

由上述递归式可以导出预测值x^*_{i+1}的多项式：

$$x_{i+1}^{*}=\sum_{j=0}^{n-1}a(1-a)^{j}x_{i-j}$$

指数平滑法例一——粮食增产预测

某市1950～1982年的粮食产量见图8—1。图中实线

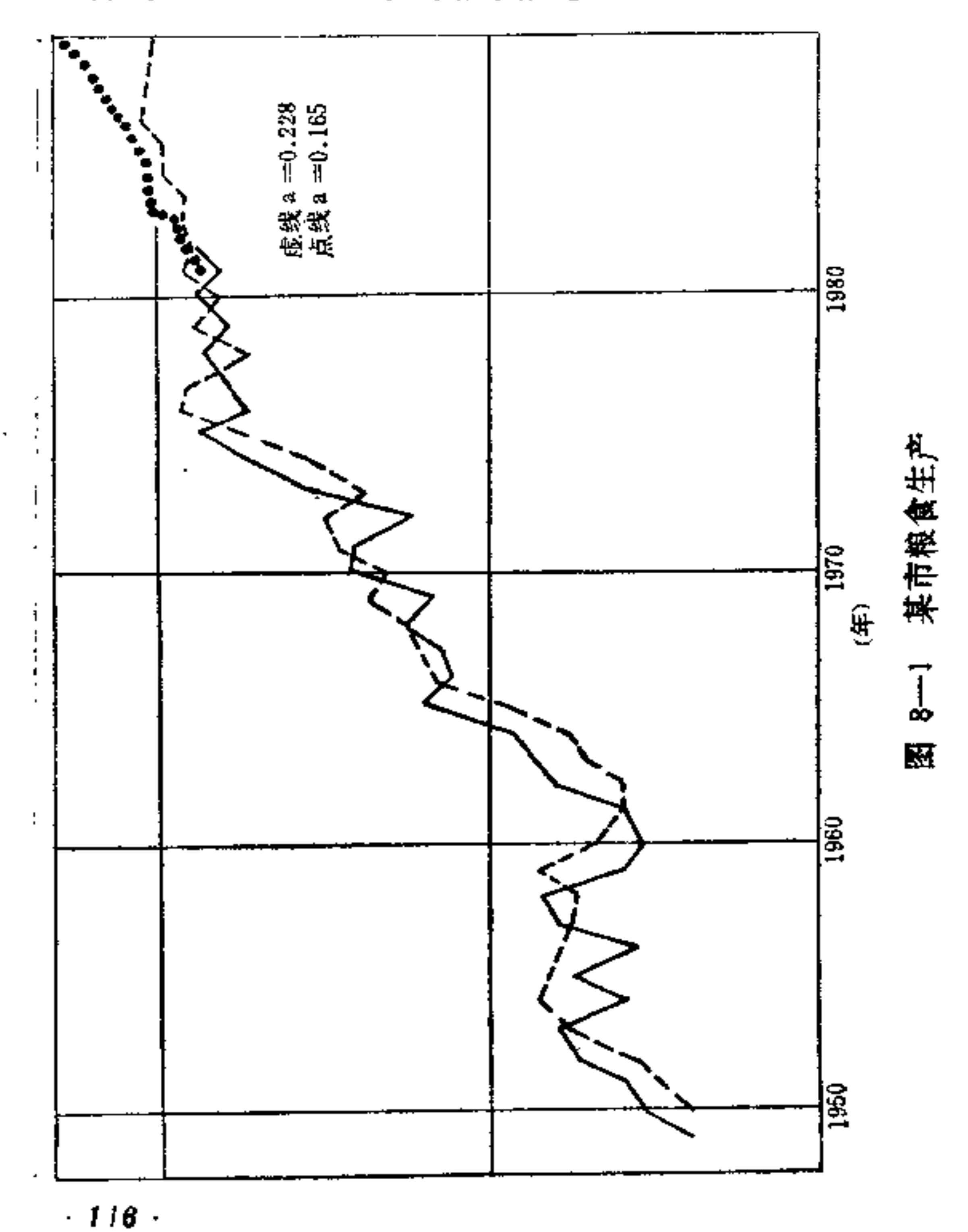


图 8—1 某市粮食生产

表示实际产量。从 1950～1964 年的 14 年，粮食翻了一番。从1964年开始预测到那一年可再翻一番呢？用指数平滑法，取a=0.228，预测值（见图8—1的虚线），预计到1987年能翻一番，但以后就增加得很慢了。如果取a=0.166，1984年就可翻一番，往后仍将继续增产（如图8—1中点线所示）。由于指数平滑法不适于远程预测，无法判定最能反映未来的预测值，但可看出，从1950年开始，用了14年粮食翻番，从1964年开始，要再翻一番，可能要用20～23年。

§7 高阶移动平均

在技术经济预测中，有时候把移动平均法中的 k 值称为移动平均的阶数，如k =1，只有上一个数据被记忆。为了使记忆较长，可以提高移动平均的阶数，并分别赋于加权系数。k阶移动平均的预测值如下：

$$x_{i+1}^{*}=M_i+a_0(x_i-x_{i-1})+a_1(x_{i-1}-x_{i-2})+\cdots+a_k(x_{i-k}-x_{i-k-1})$$

式中a_0，a_1，…，a_k是加权系数。为了使过去的记忆逐步淡漠，可以令加权系数逆时序递减[1]。

① 李卓立，《实用经济计量模型与经济预测》，清华大学出版社，59—63页，1981。

§8 混合移动平均

移动平均可以和自回归、差分等预测式混合使用。τ阶自回归和k阶移动平均的混合，记作ARMA(τ, k)，常用于假设的平稳过程。q阶差分，τ阶自回归和k阶移动平均的混合，记作ARIMA(q, τ, k)，可用于假设的齐次非平稳过程①。

以ARMA(τ, k)为例，有一种简单的表示式如下：

$$x_i = a_0 + \sum_{j=1}^{\tau} a_j x_{i-j} - \sum_{j=1}^{k} b_j x_{i-j}$$

式中前一个迭加式是τ阶自回归序列，后一迭加式是k阶加权移动平均的一种形式。根据上式编制成的一个程序采用人机联作方式给出预测值和检验结果。可用于预测某些商品的月销售量。

§9 直观随机误差

一次预测中的直观随机误差，记作($\hat{x}-x$)，其中$\hat{x}$是实际值，x是预测值。常常用($\hat{x}-x$)的某种统计量来估计一个模型的可能直观随机误差。并以此评价以后的预测精度。所以随机误差的估计也称为预测的预测。

① Box G.E.P and Jenkins G.M., Time series analysis, 1970.

$(\hat{x}_i - x_i)$ 可能是正值，也可能是负值。为了避免迭加中互相抵销，误差统计中，应采用平方值或取绝对值。

设实际序列是〈$\hat{x}_i$〉，对应的预测序列是：〈x_i〉，$i=1, 2, \cdots, n$。几种常见的随机误差的估计方法如下：

1）均方根误差：

$$\varepsilon(\mathrm{R.M.S}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{x}_i - x_i)^2}$$

如果用 $\frac{1}{(n-1)}$ 代替 $\frac{1}{n}$，就得到无偏的均方根误差。

2）平均绝对偏差：

$$\mathrm{MAD} = \mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{x}_i - x_i\right|$$

3）平均绝对百分比误差：

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{x}_i - x_i}{\hat{x}_i}\right|$$

§10 事件预测的置信水平

事件预测只有“发生”和“不发生”两种状态。它可以当作估计预测置信水平的高度退化而又比较具体的例子。

设在已加的一段时间（或空间）内，发生 n 次事

件，相应的预测是m次。n次事件中有r_n次在预先假定的条件下和预测符合，m次预测中有r_m次也在预先假定的条件下和实际符合。那么近似地估计不漏报和不错报的置信水平分别是：

$$(1-\alpha)\cong\frac{r_n}{n+1}$$

$$(1-\beta)\cong\frac{r_m}{m+1}$$

事件预测的置信水平例——北京暴雨预测的置信水平

根据一个探索性的模型，事后检查1959～1964年间北京市降水量在50毫米／日以上的日期x和实际日期$\hat{x}$比较如表8—2：

表8—2

年份	预测日期 x	实际日期 $\hat{x}$	年份	预测日期 x	实际日期 $\hat{x}$
1959	7.4		1961	9.21	9.28
	7.22	7.21(*)	1963	8.2	8.4(*)
	7.29	7.31(*)		8.10	8.8(*)
	8.10	8.6			8.9(*)
		8.13		8.21	
		8.18		8.19	
	9.8			10.26	
1960		7.16	1964		4.5
1961		7.16			7.21
		7.22		8.2	8.1(*)
		8.22		8.13	8.13(*)

表8—2中$|\hat{x}_i - x_i| \leqslant 7$天。标记(*)各项$|\hat{x}_i - x_i| \leqslant 2$天。

估计置信水平$(1-\alpha)$和$(1-\beta)$各值如表8—3。

将表8—3的置信水平用直线连接（见图8—2），两

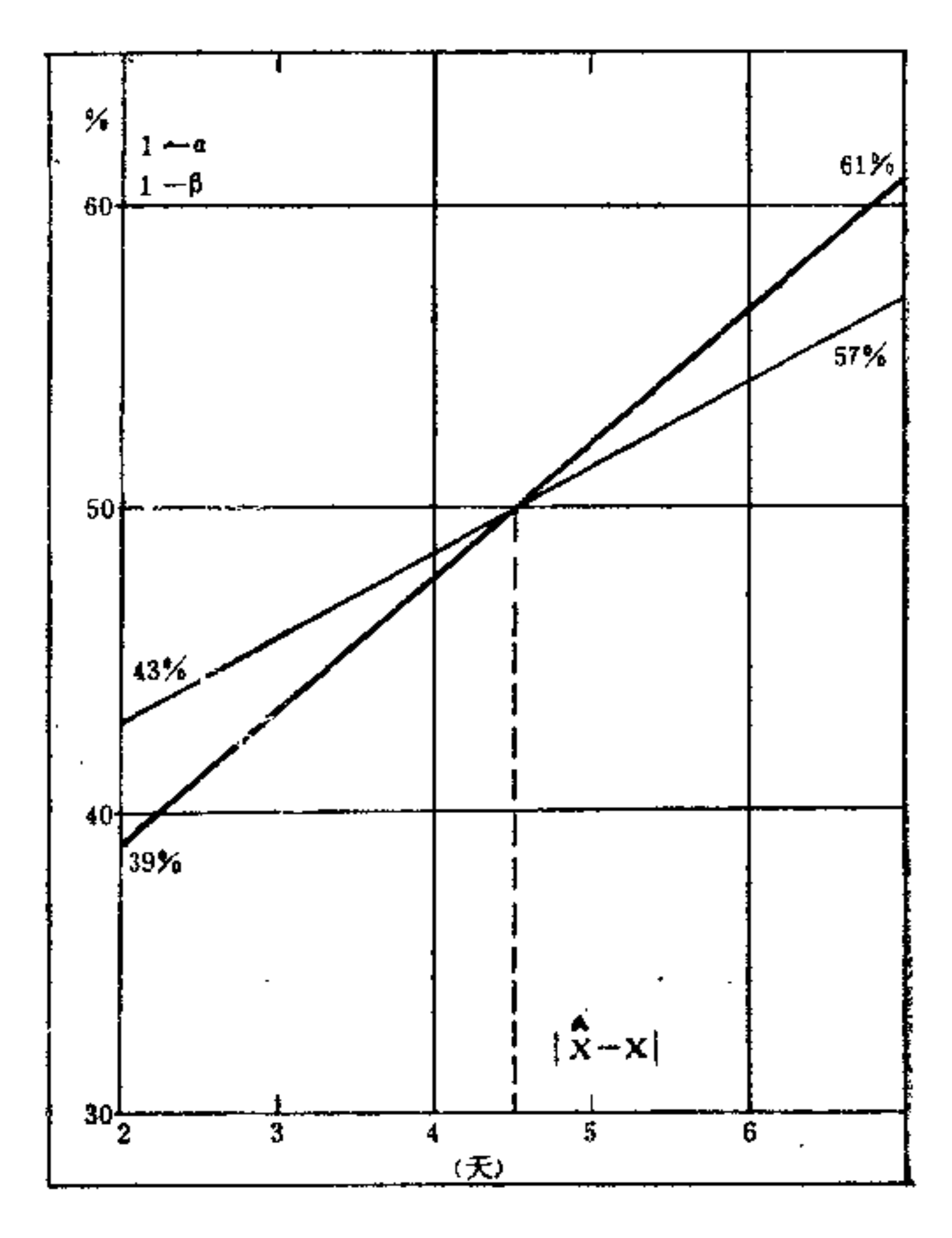


图 8—2 从两种置信水平确定可行区间

线之交点大体相当于漏报和错报的置信水平都在50%左右，预报允许错误在4～5日之间。

表8—3

$\mid \hat{x}-x \mid \leqslant$	n	m	r_n	r_m	$(1-\alpha)$	$(1-\beta)$
7天	17	13	11	8	$^{11}/_{18} \cong 61\%$	$^{8}/_{14} \cong 57\%$
2天	17	13	7	6	$^{7}/_{18} \cong 39\%$	$^{6}/_{14} \cong 43\%$

第九章　随机性的否定

习惯上随机性一词有两种用法，一种用法是相对确定性而言的，这种用法见于随机过程(stochastic process)、随机相倚、随机矩阵等。另一种用法是相对秩序而言的，这种用法见于随机事件(random event)、随机数、随机顺序、随机路线、随机过程(random process)、随机序列等。本章所说的随机性偏于后一种。

对于传播信息而言，随机性越强，信息就越弱。在某些技术领域中，把随机性成份称为干扰或噪声，把非随机性成份称为信号。如果先假设一个数据序列是完全随机性的，然后用假设检查来否定这一假设，可能取得属于信号的成份。离散型的等概率随机游动和连续型的均匀分布都是信息极贫乏的分布。

§1 均匀分布

在平稳有限域中，数据容量有限的数代数系内，数值以固定概率独立出现，这样的体系称为均匀体系。均匀分布是均匀体系内随机数的分布。设随机数在连续域

〔a，b〕均匀分布，其分布函数是：

$$F(x)=\begin{cases}1 & b<x\\ \dfrac{x-a}{b-a} & a\leqslant x\leqslant b\\ 0 & x<a\end{cases}$$

概率密度函数是：

$$f(x)=\begin{cases}\dfrac{x}{b-a} & a\leqslant x\leqslant b\\ 0 & x<a,\ x>b\end{cases}$$

在域〔a，b〕内，任何一个固定宽度Δx的区间上出现x的概率也是固定的。即

$$p(x)=\frac{\Delta x}{b-a}$$

如果Δx为全域（b－a），则$p(b-a)=1$

在〔0，1〕域内，即$a=0$，$b=1$的特款中，均匀分布的数字特征为：

数学期望：$M_x=\int_0^1 x\,dx=\frac{1}{2}$

一阶中心矩：$M(x-M_x)=0$

二阶中心矩（或称方差）：

$$D_x=M(x-M_x)^2=\int_0^1\left(x-\frac{1}{2}\right)^2dx=\frac{1}{12}$$

三阶中心矩：$M(x-M_x)^3=0$

四阶中心矩：$M(x-M_x)^4=\frac{1}{80}$，
……

均匀分布模型可以用于：

1)产生随机数。随机数已广泛地用于蒙特卡洛数值估计法、随机采样方案的设计等方面。许多统计手册上都有随机数表。

2)检测信息。本文认为，在平稳体系中，均匀分布是信息最贫乏的分布。均匀分布模型中只有一个信息参量，那就是平均值或概率P，在实际数据序列中，就是一定宽度Δx区间中的频数：

$$\lambda_{\Delta x}=\frac{n}{b-a}\cdot\Delta x$$

式中n为总数据容量，〔a，b〕是均匀分布域。

如果拒绝均匀分布作为模型的假设，就说明实际数据中含有平均频数$\lambda_{\Delta x}$以外的信息。

同样地，对于任意一个离数的数据分布，都可以将它的某一个局部分布假设为均匀分布，如果拒绝假设，就说明实际数据中含有其它信息。

均匀分布的否定例——1991年某地可能水涝

在可公度系例一中，从华中某地水灾年份中列出每次水灾年份的3～4个可公度式。从这些井井有条的秩序（也就是信息）判断，这些可公度式并非偶然。如果我们不满足于用形式上的秩序来判断并非偶然的论点，还

可以引入均匀分布的假设检查法。

现在取某地水涝年份为例。由历史记载得知：十九世纪到二十世纪中，某地共有水涝年份16次[①]。采用某种检查方法可知，其中有6次表现出高置信水平的可公度关系。如表9—1。表中实际数值$\hat{x}_i$简写为x_i。

表9—1

i	x_i	三元可公度式		X	$(1-\alpha)\geqslant$
1	1827	$x_2+x_3-x_4=1827$	$x_2+x_4-x_5=1827$	6	94%
		$x_3+x_4-x_3=1827$			
2	1849	$x_1+x_4-x_5=1849$	$x_1+x_5-x_4=1849$	7	97%
		$x_3+x_5-x_6=1849$	$x_1+x_4-x_6=1849$		
3	1887	$x_1+x_4-x_2=1837$	$x_1+x_6-x_{?}=1887$	6	94%
		$x_2+x_6-x_5=1837$			
4	1909	$x_1+x_5-x_2=1909$	$x_1+x_6-x_3=1909$	6	94%
		$x_2+x_3-x_1=1969$			
5	1931	$x_2+x_4-x_1=1931$	$x_2+x_2-x_3=1931$	5	88%
		$x_4+x_4-x_3=1931$			
6	1969	$x_3+x_4-x_1=1969$	$x_3+x_6-x_2=1969$	5	88%
		$x_4+x_2-x_2=1969$			
7*（预测）	1991	$x_2+x_6-x_1=1991$	$x_4+x_5-x_2=1991$	9	99%
		$x_5+x_3-x_1=1991$	$x_4+x_4-x_1=1991$		
		$x_6+x_4-x_3=1991$			

① 中央气象局气象科学研究院，《中国近五百年旱涝分布图集》，地图出版社，1981。

表9—1中X为实际频数。由于加法满足交换律，每个三元可公度式中，如前二项数值不同频数记为2，加前二项数值相同频数记为1。$(1-\alpha)$为置信水平。

在均匀分布的假设下，可以导出三元可公度式的平均频数应是：$\lambda\cong 0.63$次。表8—4中的X显然大于λ。可以判定，这些三元公度式不是完全偶然的。因此预测：1991年该地可能再次水涝。

§2 简单随机游动

简单随机游动是指：

$$S_n(+1,\ -1)=x_1+x_2+\cdots+x_n$$

其中x_1，x_2，…，x_n是整数集$I=\{+1,\ -1\}$中以固定概率出现的、独立分布的元素。"简单"的意义是指x_1，x_2，…，x_n等只可能是+1，和−1两个数。这一简单的随机游动问题引起了大量的研究工作，并且得出了一些深奥和惊人的结论。简单随机游动可以作为许多客观现象的模型，并且显示出不同程度的近似真实性①。例如：循环事件问题是研究随机游动中预测$S_n(+1,-1)=0$的问题。又如当x_i代表一种状态，"+"代表转移，那么S_n就成为马尔科夫链。如果x_1，x_2，代表物品、零件、

① Walter L. Smith, Handbook of operations research, foundations and fundametal, Van Nostrand Reinhold Co. p325, 1978.

机器的寿命，从随机变量的迭加演算，可得出带有互褶积函数的更新论积分式。特别引起运筹学方面研究的还有“停止问题”。它的基本思想就是预测n值，使$S_n(+1,-1)$首次达到一个固定值$S\alpha$。在这类问题中，还有吸收壁问题。原来研究自由移动的粒子被容器壁吸收的问题，被扩张到一个赌局中一个局中人的破产问题。

§3 等概率简单随机游动

简单的随机游动，虽然也涉及许多预测问题，但作为提取信息而被否定的对象，还可以再简单一些。即假设出现+1和-1的概率相等，就得到等概率简单随机游动。有时也简称为随机游动。

抛掷硬币实验是一种常用的比拟。在实验中每次出现正面记作+1，出现反面记作-1。作几次后，求两种数的和，或者两种数绝对值的差，就得到S_n。

更为原始的比拟是醉汉散步。“游动”两字的原意就是散步。假设有一醉汉在一条人行道上散步，进一步或退一步的概率相同，问定了n步后，最可能在那里找到他。

有一种方法认为：从数学期望上说，他还在原地，预测式是 ：

① 李卓立，《实用经济计量模型与经济预测》，清华大学出版社，55页，1981。

$$x_{i+1}^{*}=x_i$$

所以认为x_{i+1}^{*}，x_{i+2}^{*}，…，x_{i+n}^{*}均系x_i，没有变化，但预测误差的方差随n增大。结论是到原地去找。但n大了难于找到。

另一种方法是估计$S_n(+1, -1)$的平方值：

$$S_n^2(+1,-1)=\left[\sum_{i=1}^{n}(\pm 1)\right]^2$$
$$=n(\pm 1)^2+2\Sigma[(\pm 1)(\pm 1)]$$

式中：

$$(\pm 1)\cdot(\pm 1)=\begin{cases}+1 & (+1)\cdot(+1),\ (-1)\cdot(-1)\\ -1 & (+1)\cdot(-1),\ (-1)\cdot(+1)\end{cases}$$

因为机会相等，当n值很大时

$$\Sigma[(\pm 1)\cdot(\pm 1)]\cong 0$$

所以

$$S_n^2(+1, -1)=n$$
$$S_n^{\pm}(+1,-1)=\pm\sqrt{n}$$

考虑到平方过程中，正负两侧的值机会均等，所以单侧的偏迭加值是：

$$S_n^{+}(+1,-1)=-S_n^{-}(+1,-1)=\sqrt{n}/2$$

结论是：到离原地前后$\sqrt{n}/2$步的地方去找醉汉。这种说法比前一种积极一些。

第三种方法是用二项分布的概念。设在n次迭加中正向是i次，那么负向就是（n－i）次，迭加后，正向应是（2i－n）次。我们先考虑迭加后出现在正向的情

况，即$0\leqslant 2i-n\leqslant n$. 得到 i 的范围是：$\frac{n}{2}\leqslant i\leqslant n$. 再求正向各种可能结果的数学期望，得到正向偏选加值为.

$$S_n^+(+1,-1)=\begin{cases}\frac{n!}{2^n}\sum\limits_{i=\frac{n}{2}}^{n}\frac{(2i-n)}{i!(n-i)!} & n\text{为偶数}\\ \frac{n!}{2^n}\sum\limits_{i=\frac{n+1}{2}}^{n}\frac{(2i-n)}{i!(n-i)!} & n\text{为奇数}\end{cases}$$

同理可得负向偏选加值为：

$$S_n^-(+1,-1)=-S_n^+(+1,-1)$$

当n不大时，上式可以直接求得精确解，当n很大时，吕牛顿用德莫哇佛一拉普拉期局部极限定理导出了近似解：

当 n 很大时：

$$\frac{1}{2^n}\cdot\frac{n!}{i!(n-i)!}\cong\sqrt{\frac{2}{n\pi}}e^{-\frac{(2i-n)^2}{2n}}$$

所以：

$$S_n^+(+1,-1)\cong\begin{cases}\sqrt{\frac{2}{n\pi}}\cdot\sum\limits_{i=\frac{n}{2}}^{n}(2-in)e^{-\frac{(2i-n)^2}{2n}} & n\text{为偶数}\\ \sqrt{\frac{2}{n\pi}}\cdot\sum\limits_{i=\frac{n+1}{2}}^{n}(2i-n)e^{-\frac{(2i-n)^2}{2n}} & n\text{为奇数}\end{cases}$$

可以证明：当 n 很大时

$$S_n^+(+1,-1)=\sqrt{\frac{n}{2\pi}}(1-e^{-\frac{n}{2}})$$

即：当$n\to\infty$时，

$$\lim_{n\to\infty} S_n^+(+1,-1)\to\sqrt{\frac{n}{2\pi}}\to\infty$$

也就是说：偏迭加的结果是不收敛的。

第四种方法是用蒙特卡洛法模拟。它可以验证第三种方法的计算结果。将 $S_n^+(+1,-1)$ 简写成 $S_n(+1,-1)$。图9—1是 $n=9$ 时 $|S_9(+1,-1)|$ 的模拟值，随着模拟次数的增大，模拟值轨迹向二项分布计算值逼近。

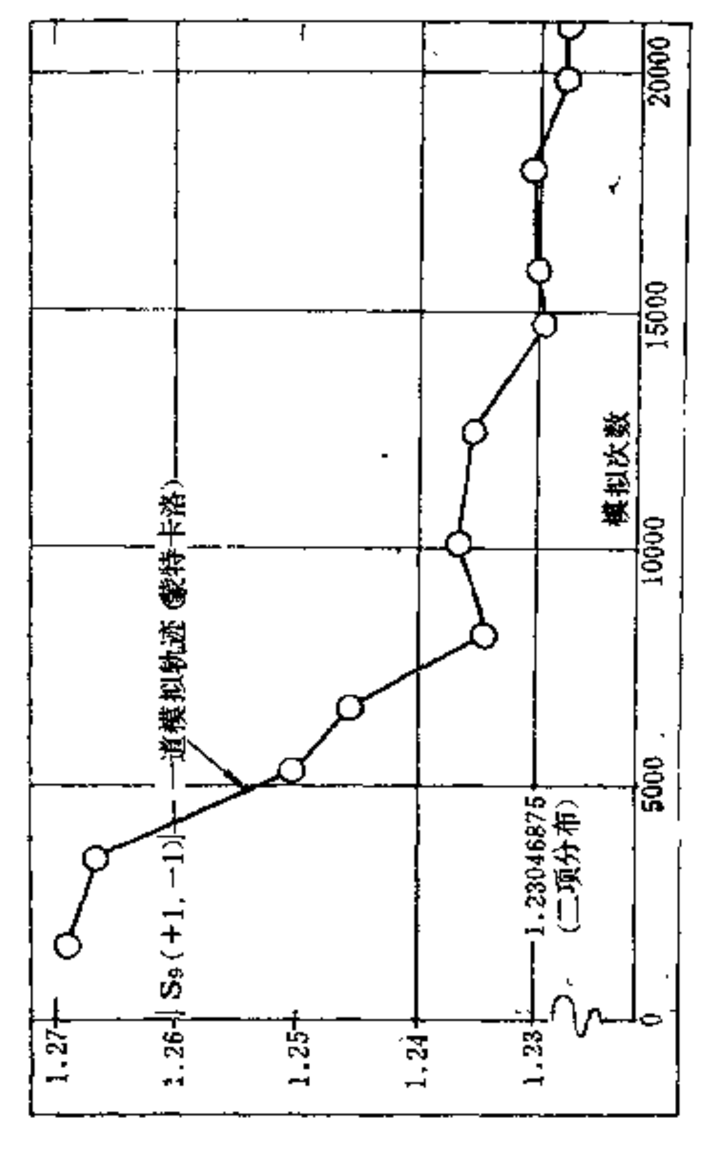


图 9—1

由于二项分布计算公式比较复杂，为了实用方便，可以采用逻辑斯谛旋回式逼近值：

$$S_n(+1, -1) = \frac{\sqrt{n/2\pi}}{1+0.234e^{-0.424\sqrt{n}}}$$

现将计算值，模拟值和逼近值列于表9—2。

表 9—2

n	$\lvert S_n(+1, -1)\rvert$		
	二项分布计算值	模拟值	逼近值
1	0.5	0.500	0.346
2	0.5	0.495	0.5
3	0.75	0.747	0.621
4	0.75	0.749	0.725
5	0.9375	0.923	0.818
6	0.9375	0.932	0.922
7	1.09375	1.093	0.981
8	1.09375	1.110	1.054
9	1.23046875	1.274	1.12
10	1.23046875	1.265	1.19
11	1.353515625	1.361	1.25
12	1.353515625	1.363	1.31

表 9—2 （续）

n	$\lvert S_n(+1,-1)\rvert$			n	$\lvert S_n(+1,-1)\rvert$		
	近似计算值	模拟值	逼近值		近似计算值	模拟值	逼近值
16	1.571		1.52	1000	12.615	13.0	12.62
20	1.762		1.72	2000	17.841	19.0	17.84
50	2.802	2.8	2.79	10^4	39.894		39.89
100	3.976	4.0	3.98	10^5	126.157		126.16
250	6.299	6.5	6.31	10^6	394.94		398.94
500	8.915	9.0	8.92	10^8	3989.42		

表9—2和表9—2（续）中，$n\geqslant12$ 用精确二项式。$16\leqslant n\leqslant20$ 用近似式。$20<n$ 用 $n\to\infty$ 极限近似式。模拟值是用蒙特卡洛法模拟，模拟次数 >2000次。逼近值用逻辑斯谛式拟合。很明显，结果是非常接近的。

$S_n(+1,-1)$ 可以作为提取信息时"纯噪音"的对应分布。当实际数据与上述结果不符，说明数据含有信息。这就是研究 $S_n(+1,-1)$ 的实用意义。

随机游动的否定例——人口中男女的分布

1982年我国人口普查结果：总人口101541万人口，男52310万人，女49231万人[①]。以男为 +1，女为 -1，得实际偏选加值对总人口的比值为：

$$\frac{\hat{S}_n(+1,-1)}{n}=\frac{3075}{101541}\cong0.030$$

① 国家统计局：《中国统计摘要，1983》，中国统计出版社，13页，1983。

如果男女人口为等概率分布，则

$$\frac{S_n(+1,-1)}{n}=\frac{1}{\sqrt{2\pi n}}\cong 0.000125$$

因此 $\hat{S}_n(+1,-1)/n \gg S_n(+1,-1)/n$

由此可见：我国人口男女分布并非等概率。

国际人口统计数字也说明，男女分布并非等概率①。研究表明：男女的自然比值从胎儿的1.2:1下降到婴儿的1.04:1，又下降到青年的1:1，最后下降到老年人的小于1:1。

随机游动的否定例二——电子中微子的质量

以随机游动的否定作为信息的定义也是随机信息模型的一个典型。仍以某种具体数据为例来说明建立这种模型的三种状态。1)纯噪音，即无信息；2)纯信息，即没有随机因素；3)带噪音的信息，即在否定噪音中取出信息。

1)纯噪音：以随机游动代表噪音。从整数集{+1,-1}中等概率地随机取n个元素。这n个元素的和称为偏迭加$S_n(+1,-1)$，数值可见表9—2。例如当$n=12$时$S_n(+1,-1)\cong 1.35$。如果我们以{cos bx}代表整数

① K.K.S, Dadzie, Scientific American 243 (3) 1980 (Sept).

集$\{+1,-1\}$,其中 bx 是随机的积,那么 n 个元素的偏迭加值为:

$$S_n(\cos bx) = \frac{2}{\pi} S_n(+1,-1)$$

当$n=12$时

$$S_n(\cos bx) \cong 0.86$$

2)纯信号:令$\cos bx_i$中的$x_i = 2\pi i$, $i=1,2,\cdots,n$。那么n个$\{\cos bx_i\}$的偏迭加之和就是b的函数:

$$S_n(\cos bx_i) = \frac{\cos n(\pi b)\sin(n+1)(\pi b)}{\sin(\pi b)} - 1$$

当$b \to 0$时,$S_n(\cos bx_i)$趋向于峰值n。例如当$n=12$时$S_n(\cos bx_i) \to 12$

3)带噪音的信息:取Au^{198}_{79}等11个β放射性同位素为例,每个元素均取β衰变中12个最强的γ射线能量作为$\{x_i\}$①,以10电子伏为单位在$1.465 \leqslant b \leqslant 1.520$变程内,以$\Delta b = 0.0001$扫描,得到6万个$S_{12}(\cos bx_i)$值,其中有48个大于8,这些异常的b值集中在$1.490 \leqslant b \leqslant 1.495$的狭区间内,平均值约为$\bar{b}=1.4925$。拒绝均匀分布的置信水平$(1-\alpha) \geqslant 90\%$。因此预计相应的能量或质量周期是$2\pi/\bar{b}$,即42.1电子伏特。因为电子中微子参与了β衰变,这一能量或质量周期应和电子中微子的静止质量有关。由此预

① 《核素常用数据表》,原子能出版社,410—411页,1977。

测；粒子物理学家将会发现电子中微子的静止质量或它的整数倍是：$M_\gamma \cong 42.1 \pm 0.07$（电子伏特）。

§4 信息的综合

来自多方面的信息需要适当综合，才能作出统一的预测。综合的基础是信息之间的关系。例如：

1)定性和定量关系：定性的判断预测和定量的计量预测本来就是技术预测的两大类别。某些体系，可以作使定性判断定量化的设计。例如，为了综合一系列事件的相互作用，客恩①设计了一种数学形式，称为影响微积分(impact calculus)。用符号表示参量间的相互影响，如“+”表示助长，“-”表示抑制，“0”表示无关等。并设计成矩阵形式，编制了专用的软件(KSIM)。

2)整体和局部的关系：如果能从整体和局部两个方面进行技术预测，然后适当综合各方面的预测结论，就可以提高预测的质量。例如：预测世界性的重要问题，如人口，能源，水源，产值等。不但要建立全球性的模型，还应建立区域性模型。

3)平行关系．有的客观体系表现出信息的多重性。在第二章的复合体系一节中，提出从一个体系可能取得

① Julius Kane, A primer for new cross-impact Language KSIM, Technological forecasting and social change 4, p129-142, 1972.

不同种类的信息。在多重体系例一中就讨论了两种信息的综合。

在单一信息体系中，也可以用不同的方法或从不同的角度处理信息，其结果未必完全相同，这就产生了加权综合问题。

4)连接关系：不同种类的信息可以有连接关系。如前项信息可以影响后项信息。在技术预测中常称为因果关系。如果不去区分因和果，那么多种信息可以构成非独立的偏相关。

5)动态关系：在多因子连接关系中，如果某项信息依照一定的时延函数影响到另一项信息，则构成动态关系。

以上这些关系也是无法彻底列举的。

信息综合的特点是：

1)主观因素占有突出地位。

2)预测程序随着结果检验不断更新，难于固定。

3)信息处理量随着综合过程迅速增加。

因此，下面给出的二个实例，只能说明对某些问题可能预测的程度，还无法详细叙述那些比较复杂并且仍在不断改进的预测过程。

信息综合例一——1983年北京天气的预测

北京天气的超远程预测是以500年来历史旱涝记载和近30年部分天气记录为依据的。在1980年预测检验的

基础上探索1983年天气概况。探索预测结论已于1982年10月25日在北京友谊宾馆休息厅中国地球物理学会全体理事会上提出。又于1982年12月7日在清华大学召开的信号处理学会上作了报告。现将预测内容和北京日报上报导的实际情况比较如表9—3。

两种不同模型的比较见图8—4。图8—4中二条折线是两种模型的计算结果，峰值预示暴雨。图中大的圆圈表示实际下暴雨日子和雨量。

单从本国历史资料中取得的综合信息，还不足以构成预报素材。如果能进一步综合全球性的和广泛区域性

表9--3

预测内容	北京日报报导内容	报导日期
"春雨降于4月4日前后"	"到昨晚（1983.4.8）11时止，近郊区雨量较大，为5～10毫米，西郊观象台达12毫米" 以后下雨日期，4月26日（20～30毫米），5月11～13日（13.3毫米），6月20日（30～40毫米）	4月9日
"夏旱"	"本市近期降雨量很少，气温又高，郊区干旱现象严重，农产受到危害。……""在一个多月的时间内，本市平均降水量则不到往年同期的百分之二十。这种雨季不雨的情况，在历史上是少见的。" "少雨高温，加速了京郊旱象的发展，对农、林、牧业生产影响很大，全市五百多万	

	亩秋粮和油料作物，都不同程度地受到干旱的影响。据有关部门粗略统计，目前受灾严重的粮田有二百多万亩，其中十五万多亩已死苗，收成无望，……”	7月23日
暴雨，“8月8日前后日降水量估计为111毫米/日。……日降水量平均误差估计为±32毫米/日。”	“本市昨（1983.8.4）降大到暴雨” “到（1983.8.4）23时止，城近郊及平原大部区县，降了大到暴雨，雨量为50到90毫米，西郊、丰台、大兴等地，降了大暴雨，雨量为110到170毫米，局部地区达200多毫米”	8月5日
	“昨晚（1983.8.5）又降大到暴雨。……城近郊区、平原各县以及密云县北部，大部分地方降雨量为40～100毫米，其中有四十多个公社达100到200毫米。大兴县的赵村、庞各庄，房山县的葫芦垡雨量最大，为200至260毫米”	8月6日
秋雨 “年景中常偏旱”……“夏旱秋涝”	“昨日（1983.8.24）夜间，我市普遍降了小到中雨，个别地区出现大雨。……”	8月25日
	“本市大部地区普降中雨”“石景山、门头沟、房山的降水量都超过30毫米，其中房山最大，达114毫米”	8月26日
	“二十五日夜，本市大部分地区又降中雨，雨量为10到25毫米”	8月27日
	“到昨晚22点为止……等地、降水量为10到15毫米。”	8月29日

的其它直接和相关信息，才可能提供可供实用的超远程预报。

信息综合例二——1983年N36°线上的线震

根据历史资料，曾于1983年5月26日向有关单位提供

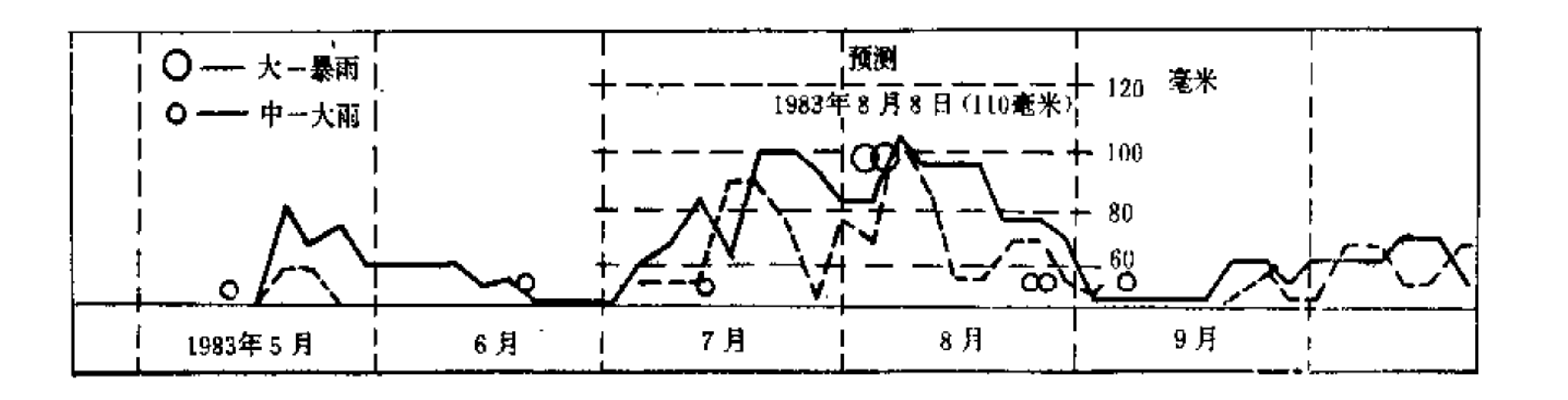


图 9-2 1983年北京暴雨预测（两种模型的综合）

"超远程预报(五)"的单项预测。结论中提出中国东部北纬36度线上可能发生第83.2号强震。地震可能日期估计为1983年9月17日。震级6.8级。大概地点："……可能在河南省新乡东西两侧北纬36°附近"。

实际地震发生于1983年11月7日，和预测日期相差51天。宏观地点在菏泽县附近，N35.2°，E115.2°，离新乡约140公里。震级差小于1级。预测和实际地震地点见图8—5。为该地人民的安全，预测至为重要。

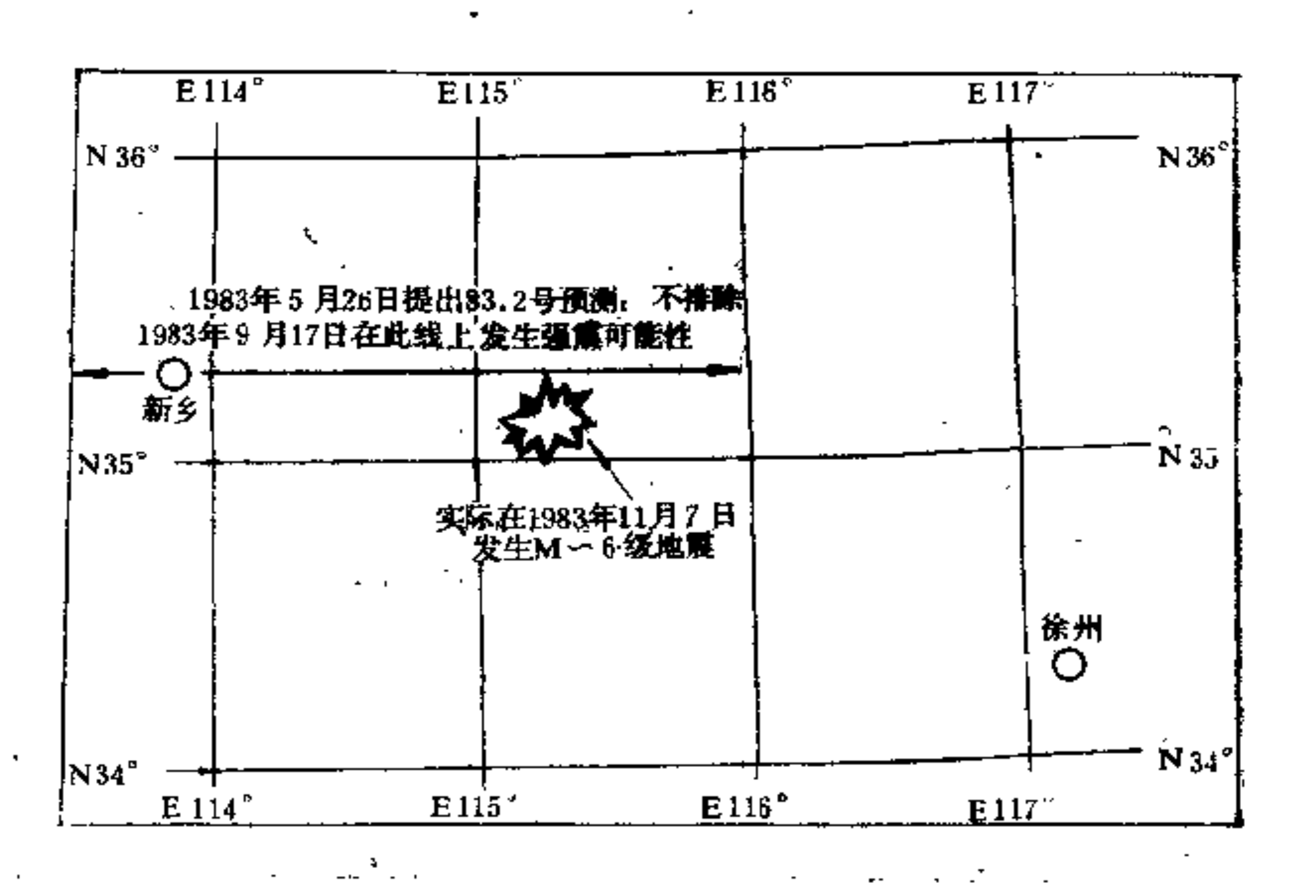


图 9—3 1983年11月7日徐州强震前预测

附　　录

泊松旋回积分式是

$$\frac{\Sigma_t Q_t}{\Sigma_\infty Q_t} = 1 - e^{-t} \sum_{i=0}^{n} \frac{t^i}{i!}$$

具体数值见下表：

t \ n	2	4	6	8	10	12	14
1	0.0803	0.0037					
2	0.3233	0.0527	0.0045	0.0002			
3	0.5768	0.1847	0.0335	0.0038			
4	0.7619	0.3712	0.1107	0.0214	0.0028		
5	0.8753	0.5595	0.2378	0.0681	0.0137	0.0020	
6	0.9380	0.7149	0.3937	0.1528	0.0426	0.0088	
7	0.9704	0.8270	0.5503	0.2709	0.1049	0.0270	0.0057
8		0.9004	0.6866	0.4075	0.1841	0.0638	0.0173
9		0.9450	0.7932	0.5393	0.2940	0.1242	0.0415
10		0.9707	0.8699	0.6672	0.4170	0.2084	0.0835
12			0.9542	0.8449	0.6527	0.4240	0.2280
14			0.9858	0.9379	0.8243	0.6415	0.4296
16				0.9780	0.9226	0.8069	0.6325
18				0.9927	0.9696	0.9083	0.7919
20						0.9610	0.8951
22							0.9523